

Verblindete Fallzahlanpassung in klinischen Studien mit Ereigniszahlendaten

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von
Simon Schneider
aus Mönchengladbach

Mannheim, 2014

Dekan: Prof. Dr. Heinz Jürgen Müller, Universität Mannheim
Referent: Prof. Dr. Martin Schlather, Universität Mannheim
Korreferent: Prof. Dr. Tim Friede, Universitätsmedizin Göttingen

Tag der mündlichen Prüfung: 25. März 2014

ZUSAMMENFASSUNG

Um die Wirksamkeit und die Sicherheit neu entwickelter Medikamente zu überprüfen, wird jährlich eine Vielzahl an klinischen Studien durchgeführt. Die vorliegende Arbeit fokussiert sich auf klinische Studien im Krankheitsbild der multiplen Sklerose, die Zähldaten als primären Endpunkt besitzen. Ereigniszähldaten werden üblicherweise durch ein Poisson- oder ein Negativ-Binomial-Modell modelliert. In der Planungsphase einer derartigen klinischen Studie ist die Fallzahlplanung ein wichtiger Baustein, um das erfolgreiche Abschließen einer Studie zu gewährleisten. Um einen Stichprobenumfang zu bestimmen, wird zunächst eine Annahme an die Kontrollrate benötigt; im Negativ-Binomial-Modell muss der Dispersionsparameter zusätzlich spezifiziert werden. Die Bestimmung der Größen, die oftmals auf historischen Daten basiert, beinhaltet eine starke Unsicherheit. Aus diesem Grund hat das Studiendesign der internen Pilotstudie in den letzten Jahren stetig an Bedeutung gewonnen. Dieses erlaubt nach Rekrutierung der ersten Patienten eine ver- bzw. entblindete Schätzung der Streuung, so dass die Fallzahl hinsichtlich ihrer initialen Annahme überprüft wird. Das Studiendesign der internen Pilotstudie bildet die Grundlage der vorliegenden Arbeit, die in drei Themenkomplexe unterteilt ist.

In einem ersten Themenkomplex wird ein analytischer Vergleich von ent- und verblindeten Verfahren zur Fallzahlanpassung im Poisson-Modell vorgestellt. Eine Bewertung der Prozeduren erfolgt mit Hilfe der entsprechenden erwarteten Fallzahl und deren Streuung zur Zwischenauswertung. Es ist festzustellen, dass das verblindete Verfahren eine systematische Verzerrung im Erwartungswert der Verteilung aufzeigt, sobald der wahre Effekt dem angenommenen Effekt nicht mehr entspricht. Demgegenüber ist die Streuung der Fallzahl bei einer verblindeten Schätzung der Störparameter in klinisch relevanten Szenarien geringer. Zusammenfassend ist festzustellen, dass verblindete und entblindete Verfahren zur Fallzahlanpassung in vielen Situationen zu geeigneten Resultaten in klinischen Studien führen. Jedoch geht die Überlegenheit einer unverblindeten Fallzahlanpassungsprozedur fast immer mit einer drastischen Unsicherheit in der Effektgröße einher, so dass andere Studiendesigns zu empfehlen sind.

Den zweiten Themenkomplex bildet eine Charakterisierung der Arbeitsweise und Leistung der vorhandenen Prozeduren im Negativ-Binomial-Modell. Es kann gezeigt werden, dass ein EM-Algorithmus

mus-basierende Ansatz zur Fallzahlanpassung in klinisch relevanten Szenarien nicht geeignet ist. Die Parameterschätzung ist abhängig von der Wahl des Konvergenzkriteriums und den Startwerten. Falls der Algorithmus nicht konvergiert, kann es zu einer Fehlinterpretation der Schätzer kommen. Demgegenüber zeigt die zusammenfassende Prozedur, die eine Gesamtrate schätzt, die gewünschten Eigenschaften hinsichtlich der Wahrscheinlichkeit, einen Fehler 1. Art zu begehen und die gewünschte Power einzuhalten. Zudem bietet die Prozedur gewisse Robustheitseigenschaften, falls die Modellannahmen im Negativ-Binomial-Modell verletzt werden.

Der dritte Themenkomplex umfasst eine Herleitung einer Methodik zur Fallzahlanpassung auf zeit-variierende Ereigniszeitdaten und deren Untersuchung in Simulationsstudien. Die entwickelte Prozedur zur Fallzahlanpassung ist unabhängig von der Wahl der (parametrischen) Baseline-Intensitätsfunktion. Des Weiteren weist das erweiterte Verfahren wünschenswerte Eigenschaften hinsichtlich der Fehlerkontrolle und Studienpower auf. Ein zusätzlicher praktischer Nutzen kann gezeigt werden, da die erweiterte Prozedur auf implementierte Routinen in Standardsoftware zurückgreift.

DANKSAGUNG

Die vorliegende Arbeit entstand im Rahmen einer dreijährigen Förderung des Projektes „Blinded sample size reestimation in clinical trials with recurrent event data and time-dependent event rates“ (FR 3070/1-1) durch die Deutsche Forschungsgemeinschaft (DFG). Für diese Förderung möchte ich mich ganz herzlich bedanken.

Mein Dank gilt Prof. Dr. Martin Schlather für die Betreuung meiner Arbeit. Für die Übernahme des Korreferats bedanke ich mich herzlich bei Prof. Dr. Tim Friede. Ich bedanke mich für die stetige Unterstützung und Förderung in jeglicher Hinsicht.

Für eine rege Unterstützung bedanke ich mich ganz herzlich bei Dr. Heinz Schmidli von der Novartis Pharma AG (Basel). Seine wertvollen Anregungen zum Thema der Arbeit haben mich nicht nur während meiner Zeit in Basel begleitet.

Weiterer Dank gilt PD Dr. Frank Konietzschke und Dr. Christian Röver für ihre ständige Hilfsbereitschaft und die kritische Durchsicht meiner Dissertation.

Meiner Schwester Christiane danke ich für das sprachliche Korrekturlesen meiner Arbeit.

Für eine unvergessene Arbeitsatmosphäre danke ich meinen Freunden, die mich auf meinem akademischen Weg begleitet haben - allen voran Hajo, David, Markus und dem restlichen Institut für Medizinische Statistik.

Schließlich gilt ein besonderer Dank meinen Eltern und Großeltern, die mich in jeder Lebensphase unterstützt haben. Meiner Schwester Barbara, Mike und den drei Kleinen danke ich für viele schöne Wochenenden. Meiner Schwester Christiane und Stefan gilt ein Dank für eine tolle Zeit in Göttingen.

Ann-Kristin danke ich herzlich für alle erhaltene Unterstützung in den letzten Jahren.

INHALTSVERZEICHNIS

1	Einleitung	1
1.1	Klinische Studien und ihre Planung	1
1.2	Multiple Sklerose als Anwendungsbeispiel	4
1.3	Zielsetzung der Arbeit	6
1.4	Aufbau der Arbeit	8
2	Statistische Modelle, Hypothesentests und Fallzahlplanung für Ereigniszählraten	9
2.1	Poisson-Modell	9
2.1.1	Statistisches Modell	9
2.1.2	Poisson-Regression	10
2.1.3	Schätzer und Hypothesentests in der Poisson-Regression	11
2.2	Negativ-Binomial-Modell	14
2.2.1	Statistisches Modell	14
2.2.2	Negativ-Binomial-Regression	15
2.2.3	Schätzer und Hypothesentests in der Negativ-Binomial-Regression	16
2.3	Weitere Möglichkeiten für die Modellierung überdispensierter Daten	18
2.4	Fallzahlplanung für Zählraten	19
2.4.1	Konzept und Bedeutung der Fallzahlplanung	19
2.4.2	Fallzahlplanung im Poisson-Modell	19
2.4.3	Fallzahlplanung im Negativ-Binomial-Modell	21
2.4.4	Fallzahlplanung für den Quasi-Likelihood Ansatz	21
3	Konzept der Fallzahlanpassung und bestehende Strategien für Zählraten	23
3.1	Studiendesigns mit Fallzahlanpassung	23
3.2	Fallzahladjustierung im Poisson-Modell	25
3.2.1	Verblindete Ansätze zur Fallzahlanpassung	25
3.2.2	Unverblindeter Ansatz zur Fallzahlanpassung	29

3.3	Fallzahladjustierung im Negativ-Binomial-Modell	29
3.3.1	Verblindete Ansätze zur Fallzahlenpassung	29
3.3.2	Unverblindeter Ansatz zur Fallzahlenpassung	31
3.4	Fallzahladjustierung für den Quasi-Likelihood Ansatz	32
3.4.1	Verblindeter Ansatz zur Fallzahlenpassung	32
3.4.2	Unverblindeter Ansatz zur Fallzahlenpassung	32
3.5	EM-Algorithmus-basierender Ansatz zur verblindeten Fallzahlenpassung	32
3.5.1	Einführung	32
3.5.2	EM-Algorithmus für das Poisson-Gamma-Regressionsmodell	33
4	Robustheitsanalyse verblindeter und unverblindeter Fallzahlenpassungsstrategien	37
4.1	Vergleich der Prozeduren im Poisson-Modell	37
4.1.1	Einleitung	37
4.1.2	Fehler 1. Art und Powervergleich	38
4.1.3	Vergleich der Fallzahlverteilungen	40
4.1.4	Beispiel	44
4.2	Robustheitsanalyse im Negativ-Binomial-Modell	48
4.2.1	Einleitung	48
4.2.2	Szenarien der Simulationsstudie	49
4.2.3	Fehler 1. Art und Powervergleich	50
4.2.4	Beispiel	54
4.2.5	Robustheitsanalyse	56
5	Erweiterung der Prozedur für zeitvariierende Raten	59
5.1	Zählprozesse	59
5.1.1	Notation und mathematische Grundlagen	59
5.1.2	Poisson-Prozess	60
5.2	Statistisches Modell, Hypothesentest und Fallzahlplanung	61
5.2.1	Statistisches Modell	61
5.2.2	Regressionsmodell	62
5.2.3	Schätzer, Hypothesentest und Fallzahlplanung	62
5.3	Verblindete Fallzahlenpassung bei zeitvariierenden Zählraten (BSSR-TREND)	64
5.4	Zeittrends in klinischen Studien im Krankheitsbild der schubförmigen multiplen Sklerose	66
5.5	Simulationsstudie motiviert durch Studien in der multiplen Sklerose	67
5.5.1	Power	69
5.5.2	Fehler 1. Art und Fallzahlverteilung	72
6	Zusammenfassung und Ausblick	75
6.1	Zusammenfassung	75
6.2	Limitationen der auf Störparameter basierten Anpassung	76

6.3 Ausblick und Erweiterungen	77
A Wahrscheinlichkeitstheorie	79
A.1 Konvergenzen	79
A.2 Stochastische Prozesse	80
A.3 Elementare, bedingte Wahrscheinlichkeiten	80
B Simulation von inhomogenen Poisson-Prozessen mit log-linearer Intensitätsfunktion	81
Literaturverzeichnis	83

ABBILDUNGSVERZEICHNIS

1.1	Schubrate zu Studienbeginn und Studienende von 56 klinischen Studien aus der Meta-Analyse in schubförmiger multipler Sklerose	5
1.2	Rückgang der Placebo- sowie Behandlungs-Schubrate im Studienverlauf	6
1.3	Jährliche Schubrate der eingeschlossenen Patienten in der Nord- bzw. Südhalbkugel. .	7
1.4	Das Grundkonzept der vorliegenden Arbeit	8
2.1	Wahrscheinlichkeitsfunktion der Poisson-Verteilung	10
2.2	Wahrscheinlichkeitsfunktion der Negativ-Binomial-Verteilung	15
4.1	Fehler 1. Art im Prozedurenvergleich bei Poissondaten	39
4.2	Grenzwerte d_L , an denen der mittlere quadratische Fehler der Prozeduren äquivalent ist	46
4.3	Fallzahlverteilung der ver- und entblindeten Anpassungsprozedur für verschiedene Fallzahlgrößen in der Zwischenauswertung	48
4.4	Verteilung der Störparameter ermittelt durch die ML-EM Prozedur	52
4.5	Simulierte Power, Fallzahl und Standardabweichung für verschiedene Szenarien. . . .	53
4.6	Darstellung der Fallzahlen resultierend aus der Anwendung der ML-EM Prozedur für die Beispielstudien in Tabelle 4.3	55
4.7	Ergebnisse der Robustheitsanalyse	57
5.1	Beispielhafter Studienablauf für unterschiedliche Rekrutierungszeiträume.	69
5.2	Power der verblindeten Prozeduren BSSR-TREND und BSSR-CONST	71
5.3	Interaktion von Studienlänge und Rekrutierungsperiode	74

TABELLENVERZEICHNIS

4.1	Betrachtete Szenarien der Simulationsstudie	38
4.2	Merkmale der geschätzten Fallzahlverteilung im Poisson-Modell	45
4.3	Eigenschaften der motivierenden klinischen Studien	45
4.4	Betrachtete Szenarien in der Simulationsstudie im NB-Modell	49
4.5	Charakteristika der ML-Lump und ML-EM Prozedur	50
4.6	Charakteristika einer unverblindeten und ML-EM Prozedur	51
4.7	Häufigkeiten auftretender kardialen Ischämien für zwei Beispielstudien	54
4.8	Simulationsaufbau der Robustheitsanalyse	56
5.1	Verteilung von $\hat{\phi}$ bei der Anwendung der BSSR-CONST Prozedur	72
5.2	Einflussfaktoren auf die Simulationsstudie	72
5.3	Verfahrenseigenschaften der BSSR-trend Prozedur	73

KAPITEL 1

EINLEITUNG

1.1 Klinische Studien und ihre Planung

Um die Wirksamkeit und Sicherheit von neu erforschten Medikamenten, Therapieformen oder medizinischen Interventionen zu überprüfen, wird jährlich eine Vielzahl an klinischen Studien durchgeführt. Die vorliegende Arbeit thematisiert die Entwicklung von Medikamenten, die gewöhnlich in vier verschiedene klinische Phasen unterteilt wird (Shreiner, 1997).

Phase I beschreibt die erstmalige Testung einer Substanz am Menschen. Hierbei sollen pharmakokinetische und -dynamische Merkmale sowie Informationen zur Sicherheit und Toxizität an oftmals gesunden Probanden gewonnen werden. Eine Ausnahme bildet die Wissenschaft der Onkologie, in der neu entwickelte Medikamente oder Kombinationstherapien in der Regel an Patienten mit fortgeschrittener Erkrankung erprobt werden. Hintergrund ist die hohe Toxizität der Anti-Krebsmedikamente, die von leichten bis zu lebensgefährlichen Nebenwirkungen am Menschen führen können. Folglich unterliegen Phase I Studien in der Onkologie strengen Kontrollen, wobei die pharmakologischen Endpunkte der Patienten fortlaufend beobachtet werden. Diese Situation unterscheidet sich grundlegend von der Entwicklung nicht-onkologischer Medikamente, die in ihrer Toxizität limitiert sind (Friede et al., 2008, Kapitel 20).

In *Phase II* Studien stehen die Dosis-Wirkungsbeziehung (Dose Finding) und ein erster Machbarkeitsnachweis (Proof of Concept) im Fokus. Manche Pharmaunternehmen differenzieren dabei in *Phase IIa* (Dose Finding) und *Phase IIb* (Proof of Concept) Studien (Chow und Liu, 2004, Kapitel 1). Die Dosisfindung dient zur Bestimmung der Dosisverträglichkeit und speziell zur Festlegung der Dosisstufen für klinische *Phase III* Studien. In onkologischen Studien wiederum geschieht die Dosisfindung oftmals bereits in der ersten Phase, in der die Verträglichkeit und Sicherheit im Vordergrund stehen.

Phase III Studien stellen konfirmatorische kontrollierte Studien dar, die zum Nachweis der Wirksamkeit und Sicherheit im Vergleich zu Placebo oder einer Standardtherapie dienen. Sie bilden die Basis des formalen Zulassungsantrages für pharmazeutische Produkte. Klinische Studien, die nach

der Einreichung bereits unmittelbar vor der Zulassung stehen, werden als *Phase IIIb* bezeichnet. In der dritten Entwicklungsphase werden in der Regel andere primäre Endpunkte als in der zweiten Phase gewählt; ein Beispielgebiet sind hierbei klinische Studien in multipler Sklerose. Aktuelle Phase II Studien zeigen als primäre Zielvariable die Verwendung der im Magnetresonanztomographen (neu)-gemessenen Läsionen. Im Gegensatz wird in Phase III Studien die jährliche Schubrate (ARR) als primärer Endpunkt verwendet (Friede und Schmidli, 2010a).

Nach der Zulassung eines Medikamentes werden die Wirkung im Langzeitverlauf, seltene Nebenwirkungen, Vergleiche mit anderen Therapien und weitere Fragestellungen in *Phase IV* Studien überprüft.

In der vorliegenden Arbeit liegt der Fokus auf randomisierten, Placebo-kontrollierten, doppelt-verblindeten klinischen Phase II und III Studien. *Randomisierung* und *Verblindung* sind zwei grundlegende Prinzipien in der klinischen Forschung und Studiendurchführung, um die auftretenden Störgrößen gleichmäßig auf die zu untersuchenden Gruppen aufzuteilen und Verzerrungen vorzubeugen.

Der Begriff der *Randomisierung* beschreibt eine zufällige Zuteilung der in der Studie eingeschlossenen Patienten in die Behandlungsgruppen. Der zugrunde liegende Zufallsmechanismus muss dabei eine verborgene Zuweisung der Patienten gewährleisten (Allocation concealment), so dass eine Beeinflussung der Ergebnisse durch Außenstehende ausgeschlossen ist (Torgerson und Roberts, 1999). Für die Durchführung der Randomisierung existieren verschiedene Ansätze: die einfache Randomisierung (elementarster Fall „Münzwurf“), die Blockrandomisierung und die stratifizierte Randomisierung (Roberts und Torgerson, 1998). Die Randomisierung dient somit der Herstellung einer Strukturgleichheit zwischen den Behandlungsgruppen.

Nicht nur die zufällige Zuteilung der Patienten verhindert eine Verzerrung der Studienergebnisse. Zugleich sollte eine *Verblindung* der teilnehmenden Personen stattfinden, damit deren Handlungsweise nicht vom Wissen der Zuteilung abhängt (Day und Altman, 2000). Es werden verschiedene Stufen der Verblindung unterschieden. Ist lediglich der Studienteilnehmer verblindet, so wird von einfach-verblindeten Studien gesprochen. In einer doppelt-verblindeten Studie besitzen weder Patient noch Prüfarzt oder Pflegepersonal Informationen über die Zuteilung der Patienten. Hat ferner das Studienteam ebenfalls kein detailliertes Wissen über die Behandlungszugehörigkeit, wird von dreifach-blinden klinischen Studien gesprochen (Day und Altman, 2000). Ist die Behandlungszugehörigkeit allen Studienteilnehmern und Ärzten bekannt, so handelt es sich um unverblindete respektive offene Studien. Zusammengefasst sorgen die Randomisierung der Patienten und die Verblindung der beteiligten Personen für eine Vergleichbarkeit der Behandlungsarme, die zu Beginn und in der fortlaufenden klinischen Studie erwünscht ist.

In der Planungsphase einer randomisierten, Placebo-kontrollierten, doppelt-verblindeten klinischen Studie ist die *Fallzahlplanung* ein wichtiger Schritt, der das erfolgreiche Abschließen der Studie wesentlich beeinflusst. Ethische wie auch ökonomische Aspekte haben einen Einfluss auf die Bestim-

mung der Studienfallzahl:

- *Ethik*: Wie bereits erläutert, werden die Patienten, die an einer randomisierten, kontrollierten und doppelt-verblindeten Studie teilnehmen, zufällig in die geplanten Behandlungsarme eingeteilt. Das Prinzip ist rechtens, so lange keine überlegenen Therapieformen verfügbar sind (Freedman, 1987, *equipoise assumption*). Das Ziel ist es nun, die Wirksamkeit eines neuen Medikaments mit einer möglichst kleinen Anzahl an Studienteilnehmern zu zeigen, um daran anschließend den nachfolgenden Patienten verbesserte Therapieansätze zur Verfügung zu stellen. Eine höhere Belastung der Patienten während der Studienphase spricht für eine Minimalfallzahl. Wird andererseits eine zu kleine Studienpopulation bestimmt, werden möglicherweise auftretende klinisch relevante Effekte übersehen.
- *Ökonomische Faktoren*: Eine klinische Studie verbraucht in mehrerer Hinsicht eine Vielzahl an oftmals beschränkten Ressourcen, wobei eine Abhängigkeit zur Studiengröße offensichtlich ist. So beeinflusst z. B. die Größe der Studienpopulation den Arbeitsaufwand im Personalbereich und hat demzufolge einen direkten Einfluss auf die Budgetplanung der Studie. Des Weiteren soll aus Patientensicht die angemessenste Therapie möglichst schnell zugänglich sein, während aus Unternehmensicht der Wettbewerb im Pharmaziebereich zählt. In diesem Zusammenhang spielt die Produkteinführungszeit (time-to-market), die zweifellos möglichst kurz sein soll, eine entscheidende Rolle. Jedoch darf auch aus ökonomischer Sicht nicht vergessen werden, dass eine zu kleine Fallzahl, die einen klinisch relevanten Effekt nicht aufdecken kann, für das Unternehmen schädlich ist.

Um beide Perspektiven hinreichend gut zu berücksichtigen, wird eine minimale Fallzahl bestimmt, die bei Vorliegen eines vorgegebenen klinisch relevanten Effekts einen statistisch signifikanten Unterschied zu einer gewünschten Wahrscheinlichkeit (Power) aufdeckt. Infolgedessen muss in der Planungsphase einer klinischen Studie ein klinisch relevanter Effekt und zusätzlich die gewünschte Power spezifiziert werden. Gängige Werte für die Power liegen bei 80 % bzw. 90 %, während die Festlegung des klinisch relevanten Effekts oftmals schwieriger ist (vgl. Abschnitt 2.4). Darüber hinaus ist die Streuung der Messwerte des primären Endpunktes ein wesentlicher Einflussfaktor für die Fallzahlberechnung. Eine Annahme an die Streuung der Messwerte ist oftmals schwierig zu treffen, auch wenn auf historische Daten zurückgegriffen werden kann. Unterschiedliche Designmerkmale (z. B. Ein- und Ausschlusskriterien) der historischen klinischen Studien tragen dazu bei, dass die Verwendung der extrahierten Schätzer zu einer unsicheren initialen Fallzahl führt. Die Planung einer Studie stellt somit häufig einen erheblichen Unsicherheitsfaktor dar. Aus diesem Grund haben Studiendesigns, die eine Fallzahlanpassung im Studienverlauf erlauben, in den letzten Jahren stetig an Bedeutung gewonnen. Eine Vielzahl an Publikationen unterstreicht den wachsenden Stellenwert der adaptiven Designs. In der vorliegenden Arbeit wird das Studiendesign der internen Pilotstudie von Wittes und Brittain (1990) betrachtet. Dieses erlaubt nach Rekrutierung der ersten Patienten eine ver- bzw. entblindete Schätzung der Streuung, so dass die Fallzahl hinsichtlich ihrer initialen

Annahme überprüft wird. Andere Studiendesigns mit Fallzahlanpassung schätzen den Effekt und können zu einem frühzeitigen Studienende führen. Eine Übersicht existierender Studiendesigns mit Fallzahlanpassung ist in Kapitel 3 dargestellt. Verblindete sowie entblindete Verfahren zur Fallzahlanpassung im Design der internen Pilotstudie sind für eine Vielzahl an Endpunkten in der Literatur zu finden. Einen Überblick über die verschiedenen Ansätze geben Friede und Kieser (2006).

Die vorliegende Arbeit beschäftigt sich mit Fallzahlanpassungsstrategien in klinischen Studien, die wiederkehrende Ereignisse als primären Endpunkt besitzen. Friede und Schmidli (2010a) beschreiben eine Strategie zur Fallzahladjustierung für Poisson-verteilte Daten; Strategien für Negativ-Binomial-verteilte Zufallsvariablen sind in Cook et al. (2009) sowie in Friede und Schmidli (2010b) gegeben. Die Wahl der genannten Verteilungsfunktionen zur Modellierung des primären Endpunktes ist in vielen klinischen Anwendungsbereichen zu beobachten. In der chronisch obstruktiven Lungenerkrankung (COPD) werden z. B. die auftretenden Exazerbationen durch Zähldaten modelliert (Keene et al., 2007, 2008, 2009). Ein weiteres Beispiel ist die schubartig verlaufende Krankheit Gicht (Akacha und Benda, 2010). Der vorliegenden Arbeit liegt die chronisch-entzündliche Entmarkungserkrankung des zentralen Nervensystems Multiple Sklerose als motivierendes Krankheitsbild zugrunde. Im folgenden Abschnitt werden klinische Studien in der multiplen Sklerose im Detail vorgestellt.

1.2 Multiple Sklerose als Anwendungsbeispiel

Eine Vielzahl der aktuellen klinischen Phase III Studien im Krankheitsbild der multiplen Sklerose verwenden als primären Endpunkt die jährliche Schubrate (ARR) oder die Anzahl der (neu-)gemessenen Läsionen, die mit Hilfe der Magnetresonanztomographie bestimmt werden. Diese Zähldaten werden üblicherweise durch ein Negativ-Binomial- oder ein Poisson-Modell modelliert. Als Beispielstudien seien die Arbeiten von Sormani et al. (1999), Sormani et al. (2001), Rudick et al. (2006), van den Elskamp et al. (2009) und Wang et al. (2009), angeführt.

Zunächst wird der Frage nach dem praktischen Nutzen der Verwendung einer internen Pilotstudie nachgegangen. In einer Meta-Analyse in der schubförmigen multiplen Sklerose wurde der Placebo-Schubratenverlauf von 56 randomisierten, Placebo-kontrollierten, doppelt-verblindeten klinischen Studien untersucht (Steinvorth et al., 2013). Abbildung 1.1 zeigt die Placebo-Schubrate vor Studienbeginn und im Studienverlauf abgetragen gegen das Publikationsjahr. Über die Zeit ist eine signifikante Reduzierung der Schubrate um 4.5 % pro Jahr zu beobachten. Ein ähnliches Phänomen wurde in früheren Arbeiten von Inusah et al. (2010) und Nicholas et al. (2011) beobachtet, wobei die der Analyse zugrunde liegenden Arbeiten nur zu einem gewissen Anteil eine Überschneidung aufweisen. Die absinkenden Placebo-Schubraten am Studienende können durch veränderte Studien-Designmerkmale (z. B. Ein- und Ausschlusskriterien, Studienlänge), dem Zeitraum zur Schubratenberechnung und durch die Schubrate zum Studienbeginn, wodurch zusätzlich der Regression zur Mitte Effekt verstärkt wird, erklärt werden (Steinvorth et al., 2013).

Die Placebo-Schubrate λ_0 stellt im Poisson- sowie im Negativ-Binomial-Modell nicht nur die er-

wartete Anzahl der Ereignisse sondern auch die Streuung bzw. einen Hauptteil der Variabilität der auftretenden Ereignisse dar. Infolgedessen liegt im Kontext der initialen Fallzahlplanung eine erhebliche Unsicherheit vor, die sowohl durch die sinkenden Placebo-Schubraten als auch durch die beachtliche Streuung der Raten innerhalb des Jahres begründet ist (vgl. Abbildung 1.1). Die Verwendung einer internen Pilotstudie, die eine Fallzahlanpassung ermöglicht, stellt folglich ein geeignetes Studiendesign dar.

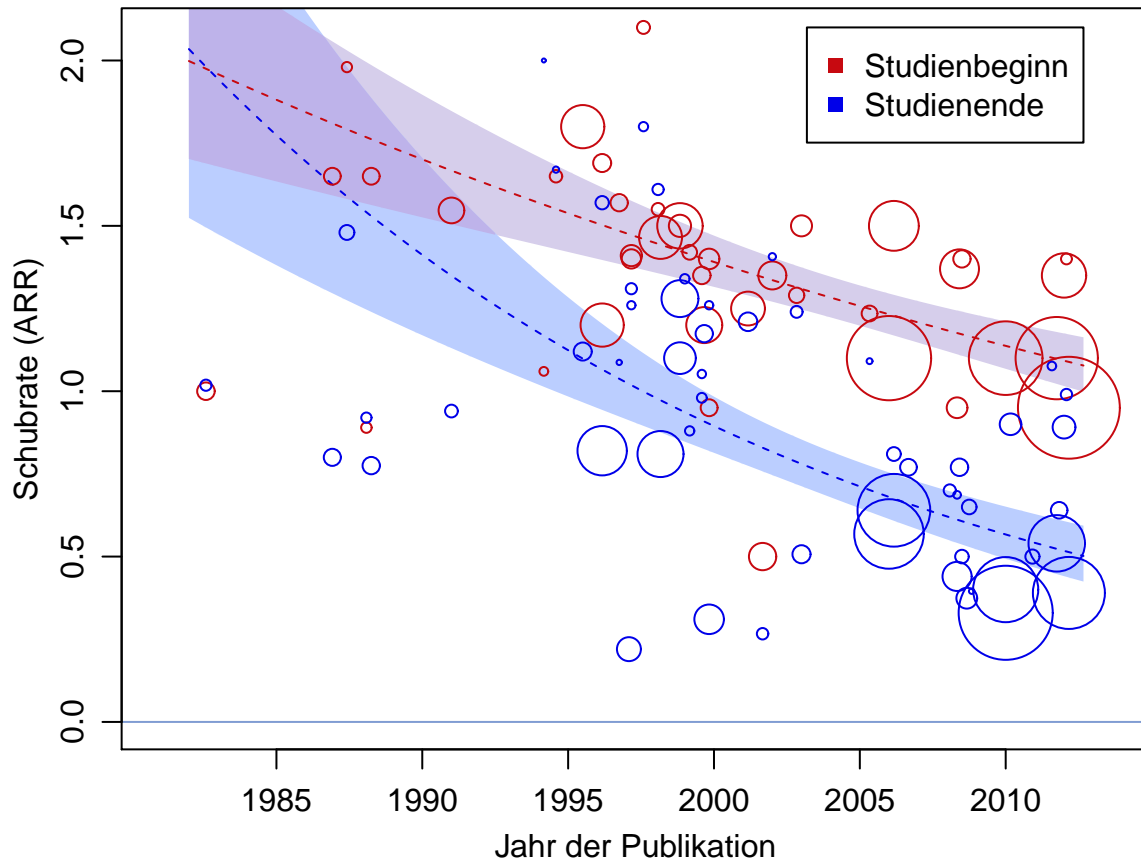


Abbildung 1.1: Schubrate zu Beginn und am Studienende von 56 randomisierten, Placebo-kontrollierten, doppelt-verblindeten klinischen Studien aufgetragen gegen das Publikationsjahr. Die Studiengröße ist durch die Größe der Kreise gekennzeichnet und ist proportional zum Inversen des Standardfehlers des Schätzers der Schubrate. Die gestrichelten Linien stellen das 95 %-Konfidenzintervall dar; verändert nach Steinvorth et al. (2013).

Eine weitere wichtige Erkenntnis wurde durch die Folgearbeit von Nicholas et al. (2012) gewonnen. Der Meta-Analyse im Krankheitsbild der schubförmigen multiplen Sklerose liegen 13 klinische Studien zugrunde. Sie zeigt als wesentliches Resultat einen Rückgang der Placebo- sowie Behandlungsrate im Verlauf der klinischen Studie. Diese Feststellung ist schematisch in Abbildung 1.2 festgehalten. Die für die Planung relevante Placebo-Rate liegt im ersten Jahr um durchschnittlich 25 % höher als im zweiten Jahr der Studie. Zu einem vergleichbaren Ergebnis gelangt Kappos et al. (2013), der in einer kürzlich erschienenen Arbeit in einer klinischen Studie einen Rückgang der jährlichen Rate gleichermaßen beobachtet (Kappos et al., 2013, Abb. 1 und Abb. 2).

Folglich ist die Annahme konstanter Raten, die den Fallzahlanpassungsmodellen von Friede und Schmidli (2010a,b) zugrunde liegen, im Falle des vorgestellten Endpunktes nicht gerechtfertigt. Wird diese zeitvariierende Komponente missachtet, können erhebliche Fehlentscheidungen entstehen. Gleiche Phänomene treten in ähnlicher Weise in anderen Krankheitsbildern auf. Wird die chronisch obstruktive Lungenerkrankung betrachtet, so folgen die Raten einem saisonalen Trend (Jenkins et al., 2012, Abb. 1.3). In den jeweiligen Wintermonaten (differenziert nach Nord- und Südhalbkugel) ist eine höhere Exazerbationsrate der Patienten zu beobachten. Diese Ergebnisse motivieren, eine Methodik herzuleiten, die bei vorliegenden zeitvariierenden Ereigniszeitdaten adäquate Ergebnisse hinsichtlich der Fallzahlplanung und Fallzahlanpassung liefert. Infolgedessen werden in einem nächsten Absatz die Ziele dieser Dissertation vorgestellt.

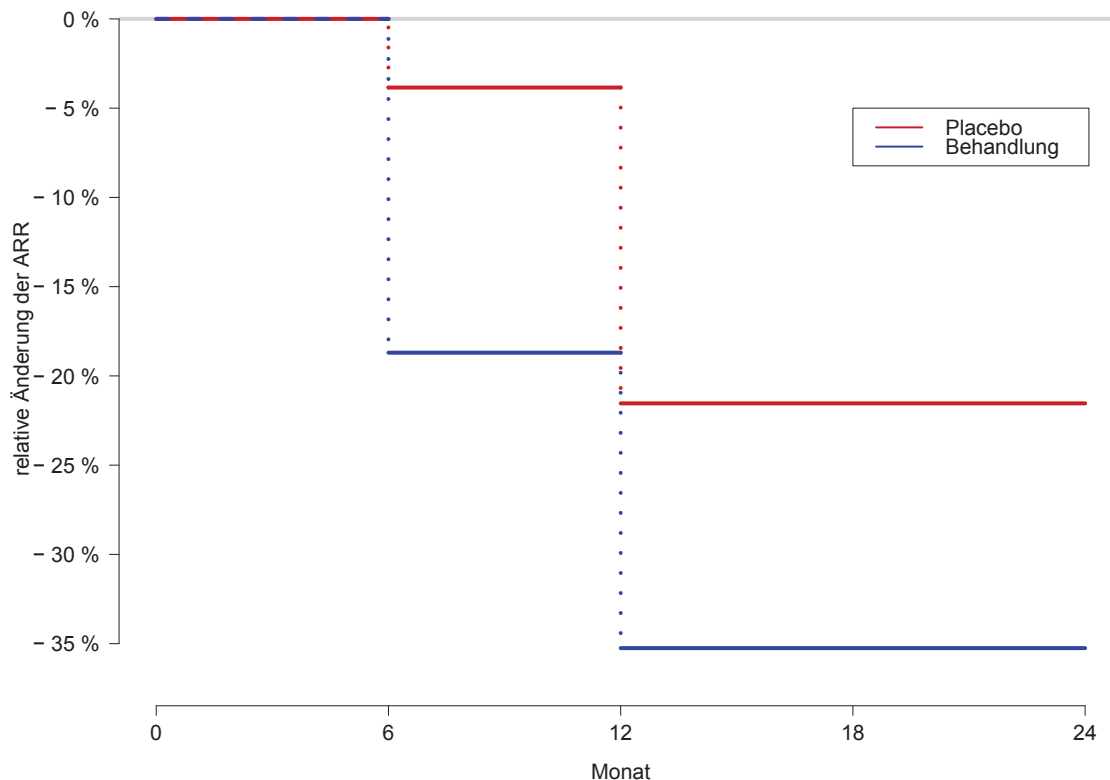


Abbildung 1.2: Rückgang der Placebo- sowie Behandlungsschubrate (%) im Studienverlauf. Die Abbildung zeigt die jährlichen Schubraten der ersten sechs Monate im Vergleich zu der zweiten Jahreshälfte (7-12 Monate) und zusätzlich zu den Monaten 13-24; verändert nach Nicholas et al. (2012).

1.3 Zielsetzung der Arbeit

Die vorliegende Arbeit beschäftigt sich mit Fallzahlanpassungsstrategien in klinischen Studien, die Ereigniszeitdaten und Zähldaten als primären Endpunkt besitzen. Die Literatur stellt Strategien zur Fallzahladjustierung für Poisson-verteilte Daten von Friede und Schmidli (2010a) sowie für Negativ-Binomial-verteilte Zufallsvariable von Cook et al. (2009) und Friede und Schmidli (2010b) bereit, die

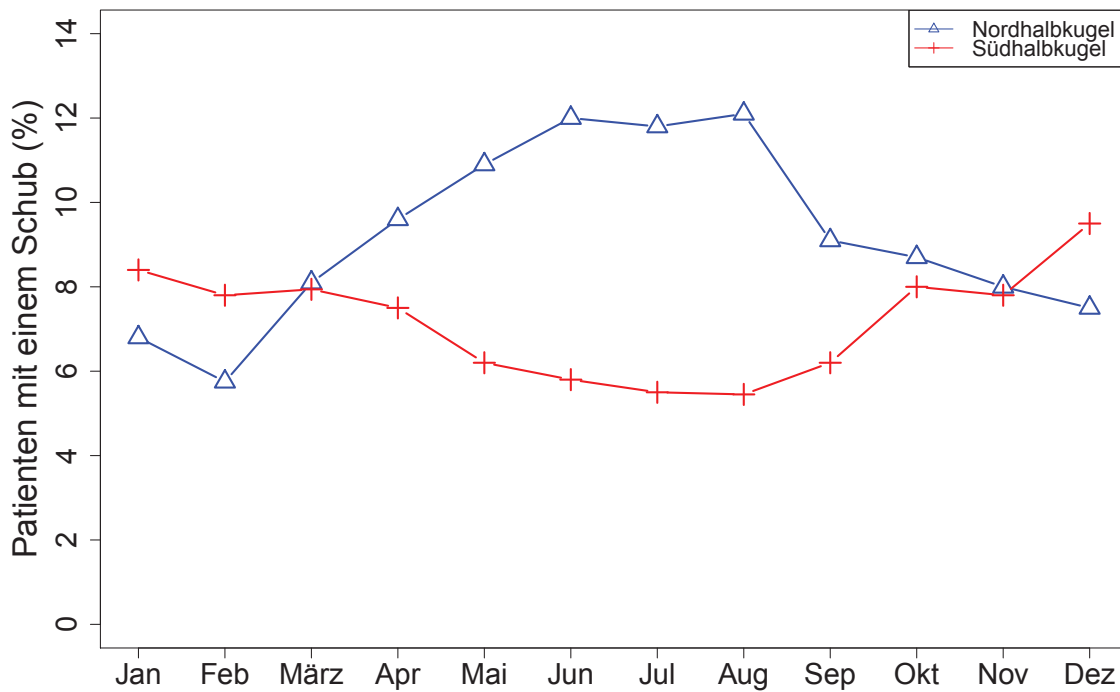


Abbildung 1.3: Jährliche Schubrate der eingeschlossenen Patienten differenziert nach Nord- bzw. Südhalbkugel; verändert nach Jenkins et al. (2012).

in den folgenden Themenkomplexen betrachtet werden.

- Ein erster Themenkomplex bildet ein analytischer Vergleich von ent- und verblindeten Verfahren zur Fallzahlanpassung im Poisson-Modell. Diesbezüglich wird in dieser Arbeit ein unverblindetes Verfahren zur Fallzahlanpassung vorgestellt. In einer detaillierten Simulationsstudie werden die Resultate überprüft.
- Den zweiten großen Themenkomplex bildet eine Charakterisierung der Arbeitsweise und Leistung der vorhandenen Prozeduren von Cook et al. (2009) und Friede und Schmidli (2010b). Diese Verfahren werden hinsichtlich der Wahrscheinlichkeit eines Fehler 1. Arts, Power und Fallzahlverteilung in Simulationsstudien untersucht. Zusätzlich erfolgt eine Robustheitsanalyse in Bezug auf die Modellannahmen.
- Wie zuvor beschrieben, existieren Ansätze von Friede und Schmidli (2010a,b), die lediglich konstante Raten zur Charakterisierung der jeweiligen Verteilungsfunktion betrachten. Demgegenüber beschreibt Cook et al. (2009) ein allgemeines Modell, in dem die Rate von der Zeit abhängt, analysiert im weiteren Verlauf allerdings den Spezialfall von konstanten Raten. Infolgedessen umfasst der dritte Themenkomplex eine Herleitung einer Methodik zur Fallzahlanpassung auf zeitvariierende Ereigniszeitdaten und deren Untersuchung in Simulationsstudien.

1.4 Aufbau der Arbeit

In Kapitel 2 werden verschiedene Modellierungsansätze für Ereigniszeitdaten, asymptotische Ergebnisse und korrespondierende Möglichkeiten zur Fallzahlplanung präsentiert. Kapitel 3 liefert bestehende wie auch neue Ansätze zur unverblindeten bzw. verblindeten Fallzahlanpassung bei Zähldaten. In Kapitel 4 wird ein Vergleich sowie eine Robustheitsanalyse der Anpassungsstrategien analytisch durchgeführt und in Simulationsstudien überprüft. Die praktische Anwendung der Ergebnisse wird in mehreren Abschnitten durch die Einbeziehung von klinischen Beispielstudien verdeutlicht.

Eine Erweiterung der bestehenden Fallzahlanpassungsstrategie von Friede und Schmidli (2010b) auf zeitvariierende Raten geschieht in Kapitel 5. Zunächst werden ein adäquates statistisches Modell, Schätzer und Möglichkeiten zur Fallzahlplanung vorgestellt. Im Anschluss wird eine Fallzahlanpassungsstrategie entwickelt und diese an motivierenden klinischen Beispielen im Krankheitsbild der multiplen Sklerose analysiert. Die Arbeit schließt mit einer Zusammenfassung und Diskussion sowie einem Ausblick auf weitere Forschungsprojekte.

Aus Gründen der Leserlichkeit sind wichtige Sätze und Definitionen in den Anhang A verlagert. Die Simulation von inhomogenen Poisson-Prozessen, die in Kapitel 5 verwendet wird, ist Anhang B zu entnehmen. Die drei Themenkomplexe sind in Abbildung 1.4 schematisch dargestellt.

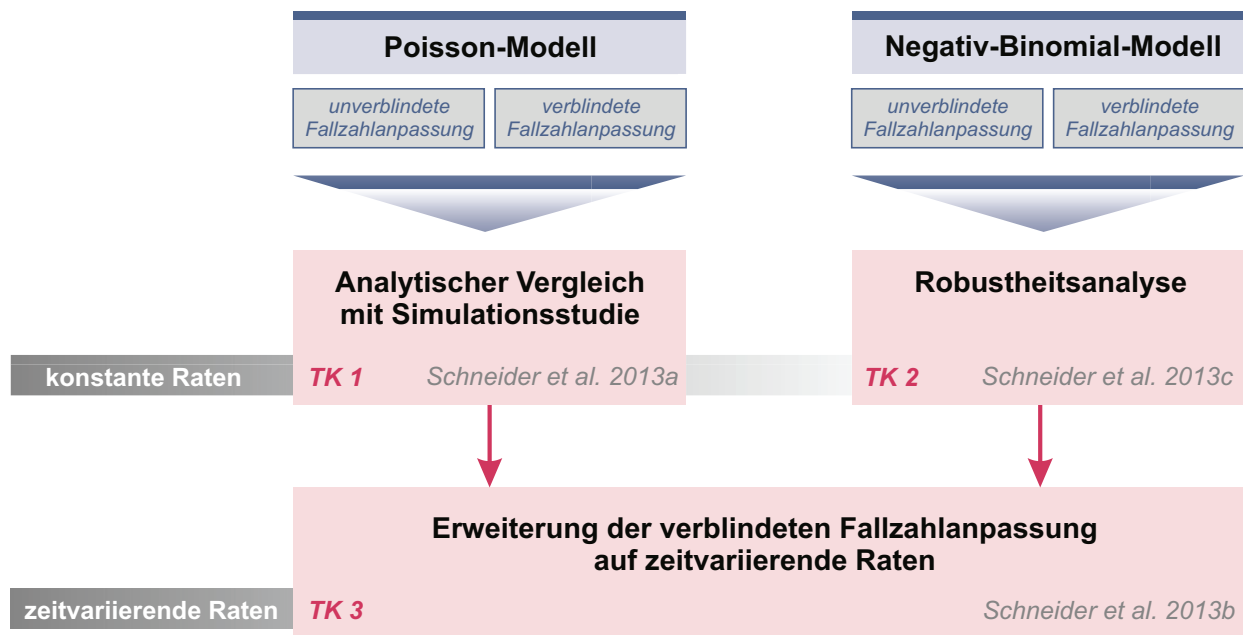


Abbildung 1.4: Das Grundkonzept der vorliegenden Arbeit.

KAPITEL 2

STATISTISCHE MODELLE, HYPOTHESENTESTS UND FALLZAHLPLANUNG FÜR EREIGNISZÄHLDATEN

In dem folgenden Kapitel werden einige Grundlagen für die Modellierung von Zähldaten gegeben. Zunächst wird das statistische Modell für Poisson-verteilte Zufallsvariablen und die daraus resultierenden Schätzer mit ihren asymptotischen Eigenschaften erläutert. Anschließend wird auf die Hypothesenformulierung und den korrespondierenden Wald-Test eingegangen. Eine Ausweitung der Ergebnisse auf Negativ-Binomial-verteilte Zufallsvariablen und eine Diskussion zur Fallzahlplanung beenden das Kapitel.

2.1 Poisson-Modell

2.1.1 Statistisches Modell

Im statistischen Modell werden unabhängig verteilte Zufallsvariablen N_{ij} mit den ganzzahligen Realisierungen $n_{ij} = 0, 1, 2, \dots$, wobei $i = 0, 1$ und $j = 1, \dots, m_i$, betrachtet. Der Index i steht im weiteren Verlauf für die Behandlungszugehörigkeit (Placebo oder Behandlung) und j für den j -ten Patienten. Die Zufallsvariablen N_{ij} heißen Poisson-verteilt mit Parameter $\lambda_i > 0$ ($N_{ij} \sim \text{Pois}(\lambda_i)$, $i = 0, 1$; $j = 1, \dots, m_i$), falls sie die folgende Wahrscheinlichkeitsfunktion besitzen:

$$p(n_{ij}) := P(N_{ij} = n_{ij}) = (n_{ij}!)^{-1} \lambda_i^{n_{ij}} \exp(-\lambda_i).$$

Eine Charakterisierung der Verteilung geschieht somit durch den Parameter λ_i , der zugleich Erwartungswert und Varianz der Verteilung ($E(N_{ij}) = \text{Var}(N_{ij}) = \lambda_i$) darstellt. Die Wahrscheinlichkeitsfunktion ist für verschiedene Parameter λ in Abbildung 2.1 exemplarisch dargestellt. Für große Ra-

ten λ konvergiert die Poisson-Verteilung gegen eine Normalverteilung. Der Parameter λ kann als erwartete Anzahl der Ereignisse interpretiert werden.

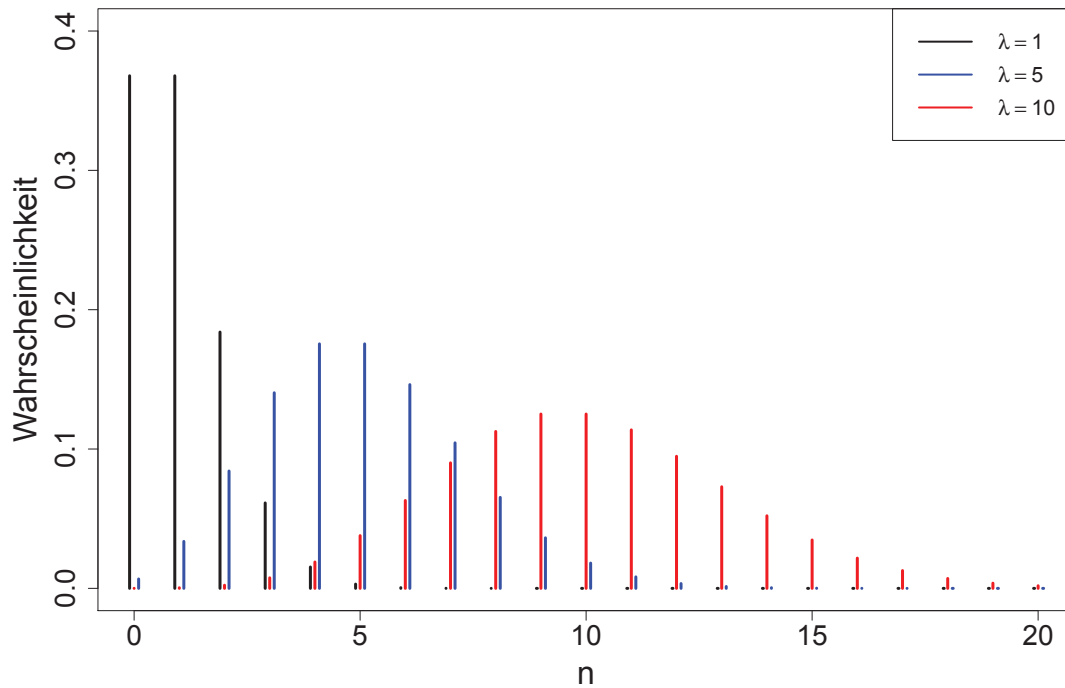


Abbildung 2.1: Wahrscheinlichkeitsfunktion der Poisson-Verteilung für verschiedene Raten λ .

Im nächsten Abschnitt wird die Poisson-Regression, die eine Modellierung der erwarteten Häufigkeiten λ_i in Abhängigkeit von fixen Kovariablen ermöglicht, vorgestellt.

2.1.2 Poisson-Regression

Zur Analyse von Poissondaten werden gewöhnlich Regressionsmodelle herangezogen. Das Ziel ist es in der Poisson-Regression einen Zusammenhang zwischen einer abhängigen Zielgröße und einem unabhängigen Kovariablenvektor \mathbb{X} herzustellen. In verallgemeinerten linearen Modellen wird dazu eine Link-Funktion $h(\cdot)$ benutzt. Der Erwartungswert λ Poisson-verteilter Daten wird üblicherweise mit einer log-lineare Linkfunktion mit dem Prädiktor verknüpft.

Im Folgenden wird das statistische Modell aus Abschnitt 2.1.1 betrachtet, wobei für jeden Patienten j in Behandlungsgruppe i eine bekannte Nachbeobachtungszeit t_{ij} in die Modellierung einfließt. Das Interesse liegt im Einfluss der Behandlung, so dass ausschließlich Regressionsmodelle mit einer fixen dichotomen Kovariablen x_i , $i = 0, 1$, betrachtet werden. Für $i = 0$, ($i = 1$) nimmt die Kovariable den Wert $x_0 = 0$, ($x_1 = 1$) an und spiegelt die Kontrollgruppenzugehörigkeit (Behandlungszugehörigkeit) des Patienten wider. Für $N_{ij}|t_{ij} \sim \text{Pois}(t_{ij}\lambda_i)$, $i = 0, 1$, $j = 1, \dots, m_i$, verteilte Zufallsvariablen wird das **Poisson-Regressionsmodell** durch

$$E(N_{ij}|x_i, t_{ij}) = t_{ij} \exp(\beta_0 + \beta_1 x_i), \quad i = 0, 1, \quad j = 1, \dots, m_i,$$

beschrieben. Die Verwendung eines log-linearen Prädiktors führt somit zu einem multiplikativen Modell, da

$$\begin{aligned} E(N_{ij}|x_i, t_{ij}) &= t_{ij} \exp(\beta_0 + \beta_1 x_i) \\ &= t_{ij} \exp(\beta_0) \exp(\beta_1 x_i), \quad i = 0, 1, \quad j = 1, \dots, m_i, \end{aligned} \quad (2.1)$$

gilt. Im weiteren Verlauf stellen $\beta_1 = \log(\lambda_1/\lambda_0)$ das logarithmierte Verhältnis der Raten (Behandlungseffekt) und β_0 die logarithmierte Kontrollrate $\log(\lambda_0)$ dar. Das führt zu einer äquivalenten Darstellung des Poisson-Regressionsmodells in Gl. 2.1 zu

$$E(N_{ij}|x_i, t_{ij}) = t_{ij} \lambda_0 \exp(\beta_1 x_i), \quad i = 0, 1, \quad j = 1, \dots, m_i.$$

Für das vorgestellte Regressionsmodell werden im nächsten Abschnitt sowohl Schätzer für die Parameter λ_0 und β_1 als auch deren asymptotische Eigenschaften untersucht. In der Regel folgt die Schätzung in generalisierten linearen Modellen mit Hilfe der Maximum-Likelihood Methode (McCulloch et al., 2001, Kapitel 5).

2.1.3 Schätzer und Hypothesentests in der Poisson-Regression

Im Zweistichprobenfall seien die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, mit Rate $t_{ij} \lambda_i$ Poisson-verteilt. Die fixe Kovariable wird wie zuvor durch x_i , $i = 0, 1$, bezeichnet; $\beta_1 = \log(\lambda_1/\lambda_0)$ stellt das logarithmierte Verhältnis der Raten dar. Mit diesen Notationen ist die Likelihood-Funktion, die als Dichte aller $M = m_0 + m_1$ Realisierungen definiert ist, durch

$$L(\lambda_0, \beta_1 | x_i, n_{ij}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} (n_{ij}!)^{-1} (t_{ij} \lambda_0 \exp(\beta_1 x_i))^{n_{ij}} \exp(-t_{ij} \lambda_0 \exp(\beta_1 x_i))$$

beschrieben. Die Likelihood wird demzufolge als Funktion von λ_0 und β_1 aufgefasst und in Abhängigkeit dieser Parameter maximiert. Da der Logarithmus eine streng monotone Transformation darstellt, führt die Maximierung der Log-Likelihood-Funktion $l(\beta_0, \beta_1 | x_i, n_{ij}) = \log(L(\lambda_0, \beta_1 | x_i, n_{ij}))$ zu äquivalenten Ergebnissen. Die Log-Likelihood-Funktion ist durch folgenden Ausdruck definiert:

$$l(\lambda_0, \beta_1 | x_i, n_{ij}) = \sum_{i=0}^1 \sum_{j=1}^{m_i} [\log((n_{ij}!)^{-1}) + n_{ij} (\log(t_{ij}) + \log(\lambda_0) + \beta_1 x_i) - t_{ij} \lambda_0 \exp(\beta_1 x_i)].$$

Die Standardvorgehensweise ist, die Nullstellen der Ableitungen von $l(\lambda_0, \beta_1 | x_i, n_{ij})$, die in den Gl. 2.2 und Gl. 2.3 gegeben sind, zu bestimmen:

$$\frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \lambda_0} = \sum_{i=0}^1 \sum_{j=1}^{m_i} [n_{ij}(1/\lambda_0) - (t_{ij} \exp(\beta_1 x_i))] \quad (2.2)$$

$$\frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \beta_1} = \sum_{i=0}^1 \sum_{j=1}^{m_i} [n_{ij} x_i - x_i t_{ij} \lambda_0 \exp(\beta_1 x_i)]. \quad (2.3)$$

Aus dem Gleichungssystem $\left(\frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \lambda_0}, \frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \beta_1} \right)' = \mathbf{0}$ ergeben sich die geschlossenen Lösungen

$$\hat{\lambda}_0 = \frac{\sum_{j=1}^{m_0} n_{0j}}{\sum_{j=1}^{m_0} t_{0j}} \quad (2.4)$$

$$\hat{\beta}_1 = \log \left(\frac{\sum_{j=0}^{m_0} t_{0j} \sum_{j=1}^{m_1} n_{1j}}{\sum_{j=1}^{m_1} t_{1j} \sum_{j=1}^{m_0} n_{0j}} \right) = \log \left(\frac{\hat{\lambda}_1}{\hat{\lambda}_0} \right), \quad (2.5)$$

wobei $\hat{\lambda}_1 = \left(\sum_{j=1}^{m_1} n_{1j} \right) / \left(\sum_{j=1}^{m_1} t_{1j} \right)$. Um asymptotische Resultate herleiten zu können, werden zunächst die folgenden Voraussetzungen an die Stichprobenumfänge und Nachbeobachtungszeiten benötigt.

Voraussetzung 2.1

- $t_{ij} \in [L, U]$, wobei $0 < L < U < \infty$.
- $M = m_0 + m_1 \rightarrow \infty$, derart, dass für $i = 0, 1 : m_i/M \rightarrow \tau_i \in (0, 1)$.
- $T \rightarrow \infty$, so dass $T_i/T \rightarrow \kappa_i \in (0, 1)$, wobei $T_i = \sum_{j=1}^{m_i} t_{ij}$ und $T = T_0 + T_1$

Mit der Voraussetzung 2.1 folgt die Konsistenz der ML-Schätzer durch Satz 2.1:

Satz 2.1 (Konsistenz I) Die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, seien mit Erwartungswert $t_{ij}\lambda_i$ unabhängig Poisson-verteilt. Dann sind die ML-Schätzer $\hat{\lambda}_0$ und $\hat{\beta}_1$ in Gl. 2.4 und Gl. 2.5 mit der Voraussetzung 2.1, im Sinne $\hat{\lambda}_0 \xrightarrow{\mathcal{P}} \lambda_0$ bzw. $\hat{\beta}_1 \xrightarrow{\mathcal{P}} \beta_1$, konsistent.

BEWEIS. Die Behauptung folgt mit Theorem 3 von Nordberg (1980), der die Eigenschaft für ein allgemeines Poisson-Modell mit p Kovariablen herleitet. Der vorliegende Fall ist lediglich ein Spezialfall. \square

Die asymptotische Normalität von $\sqrt{\frac{T_0 T_1}{T}} \left[\begin{pmatrix} \hat{\lambda}_0 \\ \hat{\beta}_1 \end{pmatrix} - \begin{pmatrix} \lambda_0 \\ \beta_1 \end{pmatrix} \right]$ wird in Satz 2.2 wie folgt gezeigt.

Satz 2.2 (Asymptotische Normalität I) Die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, seien mit Rate $t_{ij}\lambda_i$ unabhängig Poisson-verteilt. Unter der Voraussetzung 2.1 ist der Vektor $\begin{pmatrix} \hat{\lambda}_0 \\ \hat{\beta}_1 \end{pmatrix}$ asymptotisch multivariat normalverteilt mit Kovarianzmatrix Σ . Das bedeutet $\sqrt{\frac{T_0 T_1}{T}} \left[\begin{pmatrix} \hat{\lambda}_0 \\ \hat{\beta}_1 \end{pmatrix} - \begin{pmatrix} \lambda_0 \\ \beta_1 \end{pmatrix} \right] \stackrel{\mathcal{D}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$, wobei $\Sigma = \frac{T_0 T_1}{T} \mathbf{I}^{-1}$ und \mathbf{I} die Fisher-Informationsmatrix darstellt.

BEWEIS. Die asymptotische Normalität ist ein Standardresultat der ML-Theorie in Exponentialfamilien. Nordberg (1980) zeigt die asymptotische Normalität in Theorem 3 für das log-lineare Poisson-Modell. Der Beweis ist für p -Kovariablen hergeleitet und kann direkt auf den vorliegenden Spezialfall übertragen werden. \square

Für die praktische Anwendung werden Schätzer für die Varianzkomponenten gebraucht. Zunächst wird dafür die Fisher-Informationsmatrix \mathbf{I} mit Hilfe der Erwartungswerte der zweiten Ableitungen von $l(\beta_0, \beta_1 | x_i, n_{ij})$ berechnet. Die zweiten Ableitungen sind durch die Ausdrücke

$$\begin{aligned}\frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \lambda_0 \partial \lambda_0} &= \sum_{i=0}^1 \sum_{j=1}^{m_i} -n_{ij} (1/\lambda_0^2) \\ \frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \beta_1 \partial \beta_1} &= \sum_{j=1}^{m_1} -t_{1j} \lambda_0 \exp(\beta_1) \\ \frac{\partial l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial \beta_1 \partial \lambda_0} &= \sum_{j=1}^{m_1} -t_{1j} \exp(\beta_1)\end{aligned}$$

gegeben. Demnach wird die Fisher-Informationsmatrix durch $\mathbf{I} = \mathbb{E} \left[- \left(\frac{\partial^2 l(\lambda_0, \beta_1 | x_i, n_{ij})}{\partial^2 (\lambda_0, \beta_1)} \right) \right] = \begin{pmatrix} a+b & b\lambda_0 \\ b\lambda_0 & b\lambda_0^2 \end{pmatrix}$ definiert, wobei die Matrizenkomponenten durch

$$\begin{aligned}a &= \frac{\sum_{j=0}^{m_{0j}} t_{0j}}{\lambda_0} \quad \text{und} \\ b &= \frac{\sum_{j=0}^{m_{1j}} t_{1j} \exp(\beta_1)}{\lambda_0}\end{aligned}$$

gegeben sind. Daraus resultiert die inverse Fisher-Informationsmatrix zu

$$\mathbf{I}^{-1} = \begin{pmatrix} 1/a & -1/(a\lambda_0) \\ -1/(a\lambda_0) & (a+b)/(ab\lambda_0^2) \end{pmatrix} = \begin{pmatrix} \sigma_{\lambda_0}^2 & \sigma_{\lambda_0\beta_1}^2 \\ \sigma_{\lambda_0\beta_1}^2 & \sigma_{\beta_1}^2 \end{pmatrix}. \quad (2.6)$$

Damit sind die asymptotischen Resultate der Schätzer vollständig analysiert. In einem nächsten Schritt wird die Wald-Typ Statistik hergeleitet.

Um die Hypothese $H_0 : \beta_1 \geq 0$ gegen die Alternative $H_1 : \beta_1 < 0$ zu testen (einseitig), können verschiedene Teststatistiken herangezogen werden. Zu den gängigen Statistiken zählen der Score-Test, der Likelihood-Ratio Test und der Wald-Test (Cook und Lawless, 2007, Appendix A). Eine ausführliche Analyse hinsichtlich der Fehlerkontrolle und Power der Teststatistiken ist in Ng und Tang (2005) bei Poisson-verteilten und Aban et al. (2009) bei Negativ-Binomial-verteilten Daten zu finden. Die vorliegende Arbeit fokussiert den Wald-Test. Unter Verwendung der asymptotischen Normalität von $\hat{\beta}_1$, die aus Satz 2.2 hervorgeht, lässt sich die Teststatistik durch

$$TS = \sqrt{\frac{T_0 T_1}{T}} \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\hat{\sigma}_{\beta_1}^2}} \underset{\mathcal{D}}{\sim} \mathcal{N}(0, 1)$$

approximieren. Die geschätzte Varianz von $\hat{\beta}_1$ ist durch $\hat{\sigma}_{\hat{\beta}_1}^2$ gekennzeichnet. Für die Varianzberechnung von $\hat{\beta}_1$ wird die beobachtete inverse Fisher-Information an den Schätzwerten ausgewertet und verwendet, so dass sich die Teststatistik im Ausdruck

$$TS = \frac{\hat{\beta}_1}{\sqrt{\left(\frac{1}{\sum_{j=1}^{m_0} n_{0j}} + \frac{1}{\sum_{j=1}^{m_1} n_{1j}}\right)}}$$

widerspiegelt.

Mit diesen Ergebnissen sind die Grundlagen für spätere Analysen und Simulationsstudien gelegt. Jedoch weist das Poisson-Modell einige Einschränkungen in der Modellierung von Zähldaten auf. Dazu gehört die Annahme der Gleichheit von Erwartungswert und Varianz - alle Ereignisse werden somit vom gleichen Prozess bestimmt, so dass keine interindividuelle Heterogenität möglich ist. Die daraus folgende Restriktion der Gleichheit von Erwartungswert und Varianz ($E(N) = \text{Var}(N)$) ist oftmals zur Modellierung der beschriebenen Endpunkte in Kapitel 1 nicht geeignet. Somit wird im folgenden Abschnitt das Negativ-Binomial-Modell vorgestellt, das eine sogenannte Überdispersion zulässt ($\text{Var}(N) \geq E(N)$).

2.2 Negativ-Binomial-Modell

2.2.1 Statistisches Modell

In dieser Arbeit wird die Negativ-Binomial-Verteilung durch eine Poisson- und Gamma-Mischverteilung motiviert. Seien die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, mit individueller Rate $u_{ij}\lambda_i$ Poisson-verteilt, wobei u_{ij} eine Realisierung der Zufallsvariable U_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, darstellt, die einer Gamma-Verteilung mit Erwartungswert 1 und Varianz ϕ folgt. Die Dichtefunktion hat die folgende Form:

$$p(n_{ij}|u_{ij}) := P(N_{ij} = n_{ij}|U_{ij} = u_{ij}) = (n_{ij}!)^{-1} (u_{ij} \cdot \lambda_i)^{n_{ij}} \exp(-u_{ij} \cdot \lambda_i).$$

Eine marginale Verteilung von $p(n_{ij}|u_{ij})$ resultiert durch die Integration über die Gamma-verteilte Zufallsvariable U_{ij} : $\int_0^\infty p(n|u) \cdot g(u) du$. Das Resultat zeigt eine geschlossene Form der marginalen Mischverteilung, die einer Negativ-Binomial-Verteilung entspricht (Hilbe, 2007, Kapitel 8):

$$P(N_{ij} = n_{ij}|\phi) = \frac{\Gamma(n_{ij} + \phi^{-1})}{(n_{ij})! \Gamma(\phi^{-1})} \left(\frac{\phi \lambda_i}{1 + \phi \lambda_i} \right)^{n_{ij}} \left(\frac{1}{1 + \phi \lambda_i} \right)^{\phi^{-1}}.$$

Die Negativ-Binomial-Verteilung wird somit von zwei Parametern, der Rate λ und dem Shape-Parameter ϕ charakterisiert. Der Erwartungswert ist durch $E(N_{ij}) = \lambda_i$ gegeben, während die Varianz den folgenden Ausdruck besitzt: $\text{Var}(N_{ij}) = \lambda_i(1 + \phi \cdot \lambda_i)$. Der Überdispersionparameter, gegeben durch den Quotienten von Varianz und Erwartungswert, wird zu $1 + \phi \cdot \lambda_i$ berechnet. Für $\phi \rightarrow 0$ konvergiert die Verteilung gegen eine Poisson-Verteilung (Hilbe, 2007). Zur Veranschaulichung sind in Abbil-

dung 2.2 verschiedene Wahrscheinlichkeitsfunktionen der Negativ-Binomial-Verteilung dargestellt, die sich hinsichtlich ihres Shape-Parameters unterscheiden.

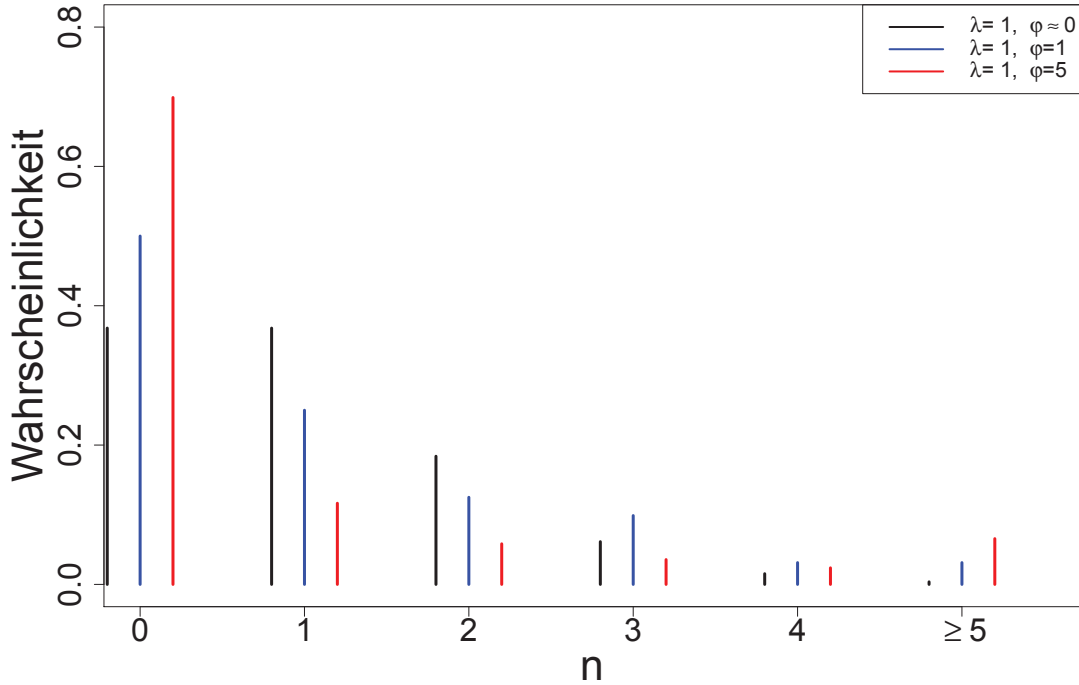


Abbildung 2.2: Wahrscheinlichkeitsfunktion der Negativ-Binomial-Verteilung für verschiedene Shape-Parameter ϕ und Rate $\lambda = 1$.

Analog zum Poisson-Modell wird im nächsten Abschnitt ein Regressionmodell vorgestellt, das eine Modellierung der erwarteten Häufigkeiten λ_i in Abhängigkeit von einer fixen Kovariablen ermöglicht.

2.2.2 Negativ-Binomial-Regression

Die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, seien mit Erwartungswert $E(N_{ij}) = t_{ij}\lambda_i$ und Varianz $Var(N_{ij}) = E(N_{ij})(1 + \phi E(N_{ij}))$ Negativ-Binomial-verteilt (geschrieben $N_{ij} \sim NB(\phi, t_{ij}\lambda_i)$). Das Interesse liegt im Einfluss der Behandlung, die durch eine fixe dichotome Kovariable x_i , $i = 0, 1$, widerspiegelt wird. Das korrespondierende marginale **Negativ-Binomial-Regressionsmodell** ist durch den Ausdruck

$$E(N_{ij}|x_i) = t_{ij}\lambda_0 \exp(\beta_1 x_i) \quad (2.7)$$

gegeben. Wird das bedingte Regressionsmodell

$$\begin{aligned}
E(N_{ij}|U_{ij} = u_{ij}, t_{ij}, x_i) &= t_{ij} u_{ij} \lambda_0 \exp(\beta_1 x_i) \\
&= t_{ij} \exp((\log(u_{ij}) + \beta_0) + \beta_1 x_i)
\end{aligned}$$

betrachtet, so ist zu erkennen, dass die modellierte Heterogenität mit den Notationen aus Abschnitt 2.1.2 als eine zufällige Verschiebung des Achsenabschnittes interpretiert werden kann. Für das vorgestellte Regressionsmodell werden im nächsten Abschnitt sowohl Schätzer für die Parameter λ_0, β_1 und φ als auch deren asymptotische Eigenschaften hergeleitet.

2.2.3 Schätzer und Hypothesentests in der Negativ-Binomial-Regression

Im Vergleich zum Poisson-Modell muss im Negativ-Binomial-Regressionsmodell ein weiterer Parameter, der Shape-Parameter φ , geschätzt werden. Mit den genannten Notationen ist die Likelihood-Funktion zum korrespondierenden Regressionsmodell in Gl. 2.7 durch

$$L(\lambda_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} \frac{\Gamma(n_{ij} + \varphi^{-1})}{n_{ij}! \Gamma(\varphi^{-1})} \left(\frac{\varphi t_{ij} \lambda_0 \exp(\beta_1 x_i)}{1 + \varphi t_{ij} \lambda_0 \exp(\beta_1 x_i)} \right)^{n_{ij}} \left(\frac{1}{1 + \varphi t_{ij} \lambda_0 \exp(\beta_1 x_i)} \right)^{\varphi^{-1}}$$

gegeben. Vergleichbar zu Abschnitt 2.2.3 wird die Log-Likelihood, die im folgenden Ausdruck gegeben ist, in Abhängigkeit der Parameter λ_0, β_1 und φ maximiert:

$$\begin{aligned}
l(\lambda_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij}) &= \sum_{i=0}^1 \sum_{j=1}^{m_i} \log(\Gamma(n_{ij} + \varphi^{-1})) - \log(\Gamma(\varphi^{-1})) \\
&\quad + n_{ij} [\log(\varphi t_{ij} \lambda_0 \exp(\beta_1 x_i)) - \log(1 + \varphi t_{ij} \lambda_0 \exp(\beta_1 x_i))] \\
&\quad - \varphi^{-1} \log(1 + \varphi t_{ij} \lambda_0 \exp(\beta_1 x_i)).
\end{aligned} \tag{2.8}$$

Die Nullstellen der Ableitungen von $l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})$ führen zu den gewünschten Schätzern. Hierbei wurde die Reparametrisierung $\beta_0 := \exp(\lambda_0)$ vorgenommen. Die ersten Ableitungen sind in den folgenden Gleichungen (Gl. 2.9, Gl. 2.10, Gl. 2.11)

$$\frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \beta_0} = \sum_{i=0}^1 \sum_{j=1}^{m_i} \frac{n_{ij} - t_{ij} \exp(\beta_0 + \beta_1 x_i)}{1 + \varphi t_{ij} \exp(\beta_0 + \beta_1 x_i)} \tag{2.9}$$

$$\frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \beta_1} = \sum_{i=0}^1 \sum_{j=1}^{m_i} \frac{x_i (n_{ij} - t_{ij} \exp(\beta_0 + \beta_1 x_i))}{1 + \varphi t_{ij} \exp(\beta_0 + \beta_1 x_i)} = \sum_{j=1}^{m_{1j}} \frac{n_{1j} - t_{1j} \exp(\beta_0 + \beta_1)}{1 + \varphi t_{1j} \exp(\beta_0 + \beta_1)} \tag{2.10}$$

$$\begin{aligned}
\frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \varphi} &= \sum_{i=0}^1 \sum_{j=1}^{m_i} \left[\left(\sum_{a=0}^{n_{ij}^*} \frac{a}{1 + \varphi a} \right) + \varphi^{-2} \log(1 + \varphi t_{ij} \exp(\beta_0 + \beta_1 x_i)) \right. \\
&\quad \left. - \frac{(n_{ij} + \varphi^{-1}) t_{ij} \exp(\beta_0 + \beta_1 x_i)}{1 + \varphi t_{ij} \exp(\beta_0 + \beta_1 x_i)} \right],
\end{aligned} \tag{2.11}$$

mit $n_{ij}^* = n_{ij} - 1$ und $\sum_{a=0}^{n_{ij}^*} a/(1 + \varphi a) = 0$ für $n_{ij} = 0$ gegeben (Lawless, 1987b). Es existieren keine geschlossenen Lösungen für das Gleichungssystem

$$\left(\frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \beta_0}, \frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \beta_1}, \frac{\partial l(\beta_0, \beta_1, \varphi | x_i, n_{ij}, t_{ij})}{\partial \varphi} \right)' = \mathbf{0},$$

so dass auf numerische Verfahren zur Nullstellenberechnung zurückgegriffen wird (z.B. Newton-Raphson Verfahren, Deuffhard (2011); Kapitel 2 ff.). Die Schätzer $\hat{\beta}_0, \hat{\beta}_1, \hat{\varphi}$ ergeben sich durch die numerische Lösung des genannten Gleichungssystems. Asymptotische Resultate werden unter der Voraussetzung 2.1 im Abschnitt 2.2.3 hergeleitet. Die Konsistenz der Schätzer folgt aus dem nachstehenden Satz.

Satz 2.3 (Konsistenz II) *Die Zufallsvariablen N_{ij} seien Negativ-Binomial-verteilt ($N_{ij} \sim \text{NB}(\varphi, t_{ij}\lambda_i)$). Dann sind die ML-Schätzer $\hat{\beta}_0, \hat{\beta}_1, \hat{\varphi}$ unter Standard-Maximum-Likelihood Argumenten und der Voraussetzung 2.1 konsistent.*

BEWEIS. Die Konsistenz der Schätzer wird in Lawless (1987a) im Appendix A gezeigt. \square

Nachfolgend wird die asymptotische Normalität der Schätzer im nächsten Satz hergeleitet.

Satz 2.4 (Asymptotische Normalität II) *Die Zufallsvariablen N_{ij} seien Negativ-Binomial-verteilt ($N_{ij} \sim \text{NB}(\varphi, t_{ij}\lambda_i)$). Unter der Voraussetzung aus Satz 2.3 sind die Schätzer $\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\varphi} \end{pmatrix}$ asymptotisch multi-*

variabel normalverteilt. Das bedeutet $\sqrt{\frac{T_0 T_1}{T}} \left[\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\varphi} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \beta_1 \\ \varphi \end{pmatrix} \right] \stackrel{\mathcal{D}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_1)$, wobei $\Sigma_1 = \frac{T_0 T_1}{T} \mathbf{I}^{-1}$ und \mathbf{I}^{-1} die Fisher-Informationsmatrix darstellt.

BEWEIS. Das Resultat der asymptotischen Normalverteilung der Schätzer ist im Appendix A in Lawless (1987a) zu finden. \square

Die Asymptotik läuft - analog zum Poisson-Modell - über die Summe der Nachbeobachtungszeiten T_0 und T_1 . Die Varianzen der Momentenschätzer von β_0 und β_1 sind durch die Einführung des Faktors $\sqrt{\frac{T_0 T_1}{T}}$ und der Voraussetzung 2.1 begrenzt, so dass diese Obergrenzen ebenfalls für die Maximum-Likelihood Schätzer aufgrund der Cramér-Rao-Ungleichung gelten (Konietschke et al., 2014). Vergleichbar zum Poisson-Modell werden für die praktische Anwendung Schätzer für die Varianzkomponenten gebraucht. Die Fisher-Informationsmatrix \mathbf{I} wird mit Hilfe der Erwartungswerte der zweiten Ableitungen bestimmt. Die inverse Matrix hat die folgende Form:

$$\mathbf{I}^{-1} = \begin{pmatrix} \mathbf{I}_{\beta_0, \beta_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma_\varphi^2 \end{pmatrix}$$

Von Interesse ist die Submatrix

$$\mathbf{I}_{\beta_0, \beta_1}^{-1} = \begin{pmatrix} a^{-1} & a^{-1} \\ a^{-1} & a^{-1} + b^{-1} \end{pmatrix} = \begin{pmatrix} \sigma_{1\beta_0}^2 & \sigma_{1\beta_0\beta_1}^2 \\ \sigma_{1\beta_0\beta_1}^2 & \sigma_{1\beta_1}^2 \end{pmatrix}, \text{ wobei}$$

$$a = \sum_{j=1}^{m_0} \frac{t_{0j} \exp(\beta_0)}{1 + \varphi t_{0j} \exp(\beta_0)} \quad \text{und} \\ b = \sum_{j=1}^{m_1} \frac{t_{1j} \exp(\beta_0 + \beta_1)}{1 + \varphi t_{1j} \exp(\beta_0 + \beta_1)}.$$

Um die Hypothese $H_0 : \beta_1 \geq 0$ gegen die Alternative $H_1 : \beta_1 < 0$ zu testen, wird ein Wald-Test herangezogen. Unter Verwendung der asymptotischen Normalität von $\hat{\beta}_1$ aus Satz 2.4 lässt sich die Teststatistik durch

$$TS_1 = \sqrt{\frac{T_0 T_1}{T}} \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\hat{\sigma}_{1\beta_1}^2}} \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1)$$

angeben, wobei $\hat{\sigma}_{1\beta_1}^2$ die geschätzte Varianz von $\hat{\beta}_1$ aus der geschätzten Fisher-Informationsmatrix $\hat{\Sigma}_1$ ist.

Für die Varianzberechnung von $\hat{\beta}_1$ wird die beobachtete inverse Fisher-Information an den Schätzwerten ausgewertet und verwendet, so dass sich die Teststatistik im Ausdruck

$$TS_1 = \frac{\hat{\beta}_1}{\sqrt{\left(\frac{1}{\sum_{j=1}^{m_{0j}} n_{0j}} + \frac{1}{\sum_{j=1}^{m_{1j}} n_{1j}} + \hat{\varphi} \left(\frac{m_{0j} + m_{1j}}{m_{0j} m_{1j}} \right) \right)}} \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1) \quad (2.12)$$

widerspiegelt.

Für den Spezialfall $t = t_{ij}$, $\forall i, j$ haben die Schätzer von λ_0 und β_1 eine äquivalente Darstellung zu den ML-Schätzern in Gl. 2.4 und Gl. 2.5 aus dem Poisson-Modell. Eine detaillierte Beschreibung der Schätzer und deren Eigenschaften ist z. B. in Zhu und Lakkis (2014) dargestellt.

2.3 Weitere Möglichkeiten für die Modellierung überdispensierter Daten

In den vorherigen Abschnitten wurden die relevanten Parameter mit Hilfe der ML-Methode geschätzt. Zur Konstruktion der Likelihood muss der probabilistische Mechanismus, durch den die Daten generiert wurden, bekannt sein. Die Spezifikation des Mechanismus gestaltet sich in vielen Szenarien als schwierig, so dass es leichter fällt, alternative Charakteristika wie den Mittelwert und die Variabilität der Zielvariablen zu beschreiben. Im Folgenden wird der sogenannte Quasi-Likelihood Ansatz vorgestellt, der lediglich Annahmen an die ersten beiden Momente der Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, stellt (Wedderburn, 1974). Die angegebene Formulierung ist absichtlich sehr allgemein gehalten und kann zur Beschreibung überdispensierter Poisson-Daten genutzt werden (McCullagh und Nelder, 1989). Seien die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, mit Erwartungswert $E(N_{ij}) = t_{ij} \lambda_i$ und Varianz $Var(N_{ij}) = \sigma^2 t_{ij} \lambda_i$ mit $\sigma^2 \geq 1$ verteilt. Vergleichbar

zu Abschnitt 2.1.2 wird die erwartete Häufigkeit in einem Regressionmodell mit einer fixen dichotomen Kovariablen x_i , $i = 0, 1$, modelliert. Eine wesentliche Erkenntnis des Quasi-Likelihood Ansatzes ist, dass die Schätzgleichungen für λ_0 und β_1 unabhängig von σ^2 sind. Demzufolge ergeben sich die gleichen Schätzer $\hat{\lambda}_0$ und $\hat{\beta}_1$ mit ihren asymptotischen Eigenschaften, die im Abschnitt 2.1.3 gegeben sind (McCullagh und Nelder, 1989). Jedoch ist die inverse Fisher-Informationsmatrix, die in Gleichung 2.6 zu finden ist, proportional zu σ^2 . Das bedeutet, dass die beobachtete Varianz in der Teststatistik mit einem Schätzer von σ^2 multipliziert werden muss. Die modifizierte Teststatistik ergibt sich somit zu

$$TS = \frac{\hat{\beta}_1}{\hat{\sigma} \sqrt{\left(\frac{1}{\sum_{j=1}^{m_0} n_{0j}} + \frac{1}{\sum_{j=1}^{m_1} n_{1j}} \right)}} \sim \mathcal{N}(0, 1).$$

Ein Momentenschätzer von σ^2 ist durch McCullagh und Nelder (1989) im allgemeinen und durch Friede und Schmidli (2010a) im speziellen Modell gegeben. Der folgende Schätzer ist eine weitere Verallgemeinerung, da er unterschiedliche Nachbeobachtungszeiten t_{ij} zulässt

$$\hat{\sigma}^2 = \left(\sum_{i=0}^1 \sum_{j=1}^{m_i} \frac{T_i}{T_i t_{ij} - t_{ij}^2} \frac{(n_{ij} - t_{ij} \hat{\lambda}_0)^2}{\hat{\lambda}_0} \right) / (m_0 + m_1), \quad (2.13)$$

wobei $T_i = \sum_{j=1}^{m_i} t_{ij}$. Zu beachten ist, dass Friede und Schmidli (2010a) den Schätzer für den Spezialfall gleicher Nachbeobachtungszeiten ($t_{ij} = t$ für alle i, j) gegeben haben. In Gleichung 2.13 ist dieser verallgemeinert. Die modifizierte Teststatistik ist nun durch $TS_{QL} = TS_P / \hat{\sigma}$ gegeben (Friede und Schmidli, 2010a; Schneider et al., 2013c).

2.4 Fallzahlplanung für Zählraten

2.4.1 Konzept und Bedeutung der Fallzahlplanung

Die Planungsphase einer randomisierten, kontrollierten und doppelt-verblindeten Studie, die den Nachweis zur Wirksamkeit eines Medikamentes bestimmt, beinhaltet die Wahl eines geeigneten Endpunktes und die Bestimmung einer adäquaten Studiengröße. Die drei wesentliche Gründe, die gegen eine zu große bzw. zu kleine Studienpopulation sprechen, wurden in der Einleitung ausführlich diskutiert (vgl. Kapitel 1). Im Folgenden werden die Fallzahlformeln für die eingeführten Modelle hergeleitet.

2.4.2 Fallzahlplanung im Poisson-Modell

Bevor der Fokus auf die Herleitung der Fallzahlformel gelegt wird, richtet sich der Blick erneut auf die Hypothese im Poisson-Modell und die korrespondierende approximativ normalverteilte Teststatistik. Es wird die Hypothese $H_0 : \beta_1 \geq 0$ gegen die Alternative $H_1 : \beta_1 < 0$ mit der korrespondierenden

$$TS = \frac{\hat{\beta}_1}{\sqrt{\left(\frac{1}{\sum_{j=1}^{m_0} n_{0j}} + \frac{1}{\sum_{j=1}^{m_1} m_{1j}}\right)}}$$

betrachtet. Im Folgenden wird davon ausgegangen, dass jeder Patient j in der Beobachtungsgruppe i die gleiche Nachbeobachtungszeit t haben soll. Somit vereinfacht sich $\sum_{j=1}^{m_0} t_{0j}$ zu $\sum_{j=1}^{m_0} t_{0j} = \sum_{j=1}^{m_0} t = m_0 \cdot t$. Zusätzlich wird von einem $1 : k$ Randomisierungsverhältnis von Kontroll- zu Behandlungsgruppe ausgegangen, so dass die Summe $\sum_{j=1}^{m_1} t_{1j}$ durch $\sum_{j=1}^{km_0} t = km_0 \cdot t$ ausgedrückt werden kann. Mit diesen Annahmen kann eine asymptotische Fallzahlformel für m_0 , die unmittelbar aus der asymptotischen Normalität der Waldtyp-Statistik folgt, bestimmt werden.

Lemma 2.1 (Fallzahlformel Poisson-Modell) *Eine asymptotische Fallzahlformel (Kontrollgruppe) für die Waldtyp-Statistik für einen angenommenen Effekt β_1^* , einer angenommenen Kontrollrate λ_0^* , einer Nachbeobachtungszeit t und einer gewünschten Power $1 - \beta$ ist durch*

$$m_0 = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\beta_1^*} \right)^2 \frac{1}{\lambda_0^* t} \left[1 + \frac{1}{k \exp(\beta_1^*)} \right]$$

gegeben, wobei z_γ das γ -Quantil der Standardnormalverteilung darstellt.

BEWEIS. Die Definition der Power folgt unmittelbar aus der asymptotischen Normalität der Teststatistik TS:

$$1 - \beta = \Phi \left(-z_{1-\alpha} - \frac{\beta_1}{\sqrt{\frac{1}{\lambda_0 m_0 t} \left[1 + \frac{1}{k \exp(\beta_1)} \right]}} \right)$$

Dieser Ausdruck vereinfacht sich zu

$$z_{1-\beta} = -z_{1-\alpha} - \frac{\beta_1}{\sqrt{\frac{1}{\lambda_0 m_0 t} \left[1 + \frac{1}{k \exp(\beta_1)} \right]}}$$

und ergibt die approximative Fallzahl für die Kontrollgruppe

$$m_0 = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\beta_1} \right)^2 \frac{1}{\lambda_0 t} \left[1 + \frac{1}{k \exp(\beta_1)} \right]. \quad (2.14)$$

□

Einige der auftretenden Größen in Gleichung 2.14 müssen in der Planungsphase einer Studie diskutiert und festgelegt werden. Die Festsetzung des Signifikanzniveaus $\alpha = 0.0125$ bzw. 0.025 oder der Power $1 - \beta = 0.8$ bzw. 0.9 stellen eine übliche Wahl dar. Hingegen wird die Frage nach der Bestimmung eines klinisch relevanten Effekts β_1^* in der Regel von Fachleuten übernommen beziehungs-

weise von Guidelines oder von einem Expertenkonsens vorgegeben.

Die Festlegung der Kontrollrate λ_0^* stellt sich im Normalfall als schwierig heraus, selbst wenn auf historische Daten zurückgegriffen werden kann (vgl. Abschnitt 1.2). Sind Studien mit dem gleichen Endpunkt verfügbar, können diese Resultate als Charakterisierung verwendet werden. Abweichungen können durch Unterschiede in Designmerkmalen, z. B. Ein- und Ausschlusskriterien der Population, entstehen. Diese Parameter werden im weiteren Verlauf als Störparameter bezeichnet. Im Abschnitt 3 werden verschiedene Studiendesigns, die eine Anpassung der Fallzahl bei Missspezifikation der Störparameter vorsehen, vorgestellt.

2.4.3 Fallzahlplanung im Negativ-Binomial-Modell

Mit Hilfe einer analogen Vorgehensweise, die in Abschnitt 2.4.2 in Lemma 2.1 benutzt wurde, wird eine approximative Fallzahlformel für das Negativ-Binomial-Modell hergeleitet. Die verwendete Fallzahlformel ist in Keene et al. (2007) zu finden und wurde von Friede und Schmidli (2010b) auf unbalancierte Stichprobenumfänge ausgeweitet:

$$m_0 = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\beta_1^*} \right)^2 \left(\frac{1}{\lambda_0^* T} \left[1 + \frac{1}{k \exp(\beta_1^*)} \right] + \frac{k+1}{k} \varphi^* \right). \quad (2.15)$$

Vergleichbar zu Absatz 2.4.2 bedarf es einer Charakterisierung der notwendigen Parameter für die Durchführung einer Fallzahlplanung. Zu der Unsicherheit der Festlegung der Kontrollrate stellt der Shape-Parameter φ^* eine weitere Herausforderung dar.

2.4.4 Fallzahlplanung für den Quasi-Likelihood Ansatz

Für den Quasi-Likelihood Ansatz kann die Fallzahlformel aus Lemma 2.1 herangezogen werden, da sich die Teststatistiken lediglich im Standardfehler um den Faktor σ unterscheiden:

$$\begin{aligned} m_0^{\text{QL}} &= m_0 \sigma^2 \\ &= \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\beta_1^*} \right)^2 \frac{1}{\lambda_0^* T} \left[1 + \frac{1}{k \exp(\beta_1^*)} \right] \sigma^2. \end{aligned} \quad (2.16)$$

Die Berechnung folgt sofort mit Hilfe von Lemma 2.1.

KAPITEL 3

KONZEPT DER FALLZAHLANPASSUNG UND BESTEHENDE STRATEGIEN FÜR ZÄHLDATEN

3.1 Studiendesigns mit Fallzahlanpassung

Die Notwendigkeit, die Störparameter - Kontrollrate und/ oder Shape-Parameter - in einer laufenden klinischen Studie zu überprüfen und gegebenenfalls anzupassen, scheint sinnvoll, wenn die auftretenden Unsicherheiten, die in Abschnitt 1.2 erläutert wurden, betrachtet werden. Es gibt verschiedene Formen und Philosophien für die Methodik und Umsetzung der Fallzahlanpassung. Zunächst wird zwischen externen und internen Pilotstudien unterschieden.

Eine externe Pilotstudie dient zur Charakterisierung der Störparameter, wobei nur eine geringe, jedoch ausreichende Anzahl an Patienten für die Schätzung der Parameter eingeschlossen wird. Die Daten der Pilotstudie werden jedoch nur zur Bestimmung einer initialen Fallzahlplanung benutzt. Danach beginnt die Rekrutierung der Studienteilnehmer, so dass die Daten der Pilotstudie nicht weiter verwendet werden. Angesichts der in Abschnitt 1.1 dargestellten ökonomischen und ethischen Aspekte weist das Vorgehen Defizite hinsichtlich der Effizienz auf. Aus diesem Grund wird in der vorliegenden Arbeit ein abweichender Ansatz verfolgt.

Es handelt sich um das Studiendesign der *internen Pilotstudie* (IPS) von Wittes und Brittain (1990), welches einen Data Review vorsieht. Das Design der internen Pilotstudie findet immer größeren Anklang in der medizinischen Forschung, da flexible Möglichkeiten zur Fallzahlanpassung eröffnet werden. Die interne Pilotstudie nach Wittes und Brittain (1990) wird über drei Komponenten beschrieben, die im Folgenden für das Poisson-Modell vorgestellt werden.

- *Initiale Fallzahlschätzung*: Die aus historischen Daten spezifizierte Kontrollrate λ_0^* wird in die approximative Fallzahlformel in Gl. 2.14 eingesetzt. Da die Fallzahl von weiteren Parametern abhängt, wird sie als Funktion derer ausgedrückt und mit $M := (k + 1) \cdot m_0(\alpha, \beta, \beta_1^*, \lambda_0^*)$

bezeichnet.

- *Fallzahlanpassung:* Nach der Rekrutierung von $M_{prop} = \pi M$, $\pi \in (0, 1]$ Patienten wird der Störparameter λ_0 , *unverblindet* oder *verblindet* aus den vorhandenen Daten geschätzt. Unverblindet bedeutet, dass die Gruppenzugehörigkeit (Kontroll- oder Behandlungsarm) der Patienten - im Gegensatz zur verblindeten Schätzung - bekannt ist. Die resultierende und möglicherweise angepasste Fallzahl wird als $\hat{M}_{re} := (k + 1) \cdot m_0(\alpha, \beta, \beta_1^*, \hat{\lambda}_{0(re)}^*)$ bezeichnet und zur Fallzahladaption verwendet. Der Index „re“ deutet auf den Fallzahlanpassungsschritt (Reestimation) hin.

Die ursprüngliche Arbeit von Wittes und Brittain (1990) beinhaltet lediglich die Möglichkeit einer Fallzahlerhöhung, was aus folgender Adaptionformel ersichtlich wird:

$$M_{(re)}^* = \max(M, \hat{M}_{re}) - M_{prop}, \quad (3.1)$$

wobei $M_{(re)}^*$ die noch zu rekrutierende Patientenzahl darstellt. Eine weitere und in dieser Arbeit verwendete Adaptionregel stammt von Birkett und Day (1994), die im Gegensatz zu Gl. 3.1 eine Korrektur der Fallzahl nach oben und unten zulässt. Die folgende Regel wird auch als *naive Adaptionregel* bezeichnet.

$$M_{(re),(naive)}^* = \max(M_{prop}, \hat{M}_{re}) - M_{prop} \quad (3.2)$$

Eine ausführliche Diskussion weiterer Adaptionstrategien ist in Friede (1998) gegeben.

- *Finale Auswertung:* Die finale Auswertung wird mit allen rekrutierten Patientendaten $M = M_{prop} + M_{(re),(naive)}^*$ durchgeführt. Der beobachtete Effekt wird mit Hilfe der Formel in Gl. 2.5 geschätzt und ein Wald-Test wird zur statistischen Testung der Hypothese verwendet.

Das Prinzip der IPS kann ebenfalls auf das Negativ-Binomial-Modell übertragen werden. Als Störparameter werden im Schritt der Fallzahlanpassung die Kontrollrate λ_0 und der Shape-Parameter φ aus den vorhandenen Daten geschätzt. Im Folgenden wird auf spezifische Begriffe wie ent- bzw. verblindete Verfahren und auf die Wahl des Parameters π eingegangen.

- *Die Wahl von π :* In verschiedenen Simulationsstudien für verschiedene Endpunkte (Wittes und Brittain, 1990; Friede et al., 2007; Friede und Schmidli, 2010a,b; Schneider et al., 2013c,a) wird die Hälfte der initial berechneten Patientenzahl $\pi = 1/2$ als IPS-Größe benutzt. Dieses Vorgehen weist gute Eigenschaften hinsichtlich der Power und der Fehlerkontrolle auf. Weitere Quellen belegen ebenfalls zufriedenstellende Ergebnisse für variierende IPS-Größen (Kieser und Friede, 2003; Friede und Schmidli, 2010a). In dieser Arbeit wird zunächst der Wert $\pi = 1/2$ verwendet. Zusätzlich wird in einem späteren Kapitel die Rekrutierungsgeschwindigkeit in Kombination mit der Studiendauer mit aktuellen Ergebnissen diskutiert.
- *Vor-/Nachteile von un- und verblindeter Schätzung:* Eine unverblindete Schätzung der Störparameter ist einfach durchzuführen, beinhaltet jedoch den Nachteil, dass ein externer Aus-

schuss (independent data monitoring committee) benötigt wird, um eine Verblindung der am Projekt beteiligten Personen zu sichern. Die Anwendung von verblindeten Verfahren erspart den Expertenausschuss. Aus methodischer Sicht sind verblindete Ansätze jedoch in der Regel von der „richtigen“ Annahme an den Effekt abhängig.

Zudem empfehlen internationale Richtlinien die Implementierung verblindeter Verfahren. So heißt es in der Richtlinie der Food and Drug Administration (2010) „Sample size adjustment using blinded methods to maintain desired study power should generally be considered for most studies“. Ein Methodenvergleich der verblindeten und entblindeten Verfahren hinsichtlich Power und der Wahrscheinlichkeit eines Fehler 1. Art wird in Kapitel 4 und in Schneider et al. (2013a) erläutert.

- *Einfluss auf den Fehler 1. Art:* Die Wahrscheinlichkeit eine vermeintliche Wirksamkeit einer unwirksamen Therapie zu finden, wird als Konsumentenrisiko bezeichnet und darf ein vordefiniertes Signifikanzniveau nicht überschreiten. Werden interne Pilotstudien mit adaptiven Fallzahlanpassungsstrategien in klinischen Studien durchgeführt, so darf es nicht zu einer Inflation des Niveaus führen. Internationale Richtlinien zur Durchführung klinischer Studien betonen diese Anforderung (ICH, 1999; Food and Drug Administration, 2010; Committee for Medicinal Products for Human Use, 2007). In den folgenden Kapiteln werden die verschiedenen Verfahren hinsichtlich dieses Gütekriteriums sowie auf praktische Relevanz untersucht.

3.2 Fallzahladjustierung im Poisson-Modell

3.2.1 Verblindete Ansätze zur Fallzahlanpassung

In diesem Abschnitt werden zwei Verfahren zur *verblindeten* Fallzahladjustierung im Poisson-Modell (Abschnitt 2.1.1) vorgestellt. Zuvor werden einige Notationen eingeführt, die zu einem besseren Verständnis beitragen. Die Anzahl der Patienten in der IPS wird mit einem Sternchen gekennzeichnet und m_0 sei zunächst die initiale Fallzahl für die Kontrollgruppe. Die Proportion der Patienten aus der Kontrollgruppe, die in die interne Pilotstudie eingeschlossen wird, ist durch $m_0^* := \pi m_0$ gegeben. Der Anteil der Patienten in der Behandlungsgruppe in der IPS wird folglich mit $m_1^* := k m_0^*$ bezeichnet. Die Variable π hat den Definitionsbereich $(0, 1]$ und steht für den Proportionsfaktor. Im weiteren Verlauf identifiziert ein Index (b) bzw. (u) ein verblindetes oder unverblindetes Verfahren. Zunächst wird die Definition der Mischverteilung zweier Wahrscheinlichkeitsfunktionen benötigt.

Definition 3.1 Seien $P_1(x)$ und $P_2(x)$ Wahrscheinlichkeitsfunktionen zweier diskreter Verteilungen und sei $0 \leq k^* \leq 1$, so heißt

$$P(x) = k^* P_1(x) + (1 - k^*) P_2(x)$$

die Wahrscheinlichkeitsfunktion der Mischverteilung von $P_1(x)$ und $P_2(x)$.

Das Randomisierungsverhältnis k kann mittels Mischparameter durch $k^* = k/(k+1)$ ausgedrückt werden. Die Mischverteilung $P(x)$ ist selbst wieder eine Wahrscheinlichkeitsfunktion, was folgender Satz sichert.

Satz 3.1 • Die Mischverteilung $P(x)$ ist selbst wieder eine Wahrscheinlichkeitsfunktion.

- Für den Erwartungswert bzgl. der Mischverteilung gilt: $E(X) = k^* \mu_1 + (1 - k^*) \mu_2$, wobei μ_i der zugehörige Erwartungswert der Wahrscheinlichkeitsfunktion $P_1(x)$ bzw. $P_2(x)$ ist.
- Seien $P_1(x)$ und $P_2(x)$ Wahrscheinlichkeitsfunktionen mit Erwartungswerten μ_i , $i = 1, 2$, und Varianzen σ_i^2 , $i = 1, 2$, dann gilt für die Varianz der Mischverteilung

$$\sigma^2 = k^* \sigma_1^2 + (1 - k^*) \sigma_2^2 + k^* (1 - k^*) (\mu_1 - \mu_2)^2.$$

BEWEIS.

- $P(x) \geq 0$ für alle x , da für die einzelnen Wahrscheinlichkeitsfunktionen $P_i(x) \geq 0$, $i = 1, 2$, für alle x gilt und die Mischungsparameter positiv sind.
- $P(x) > 0$ für höchstens abzählbar unendlich viele x , da dies für die einzelnen Wahrscheinlichkeitsfunktionen $P_1(x)$ und $P_2(x)$ gilt.
- Es gilt:

$$\begin{aligned} \sum_x P(x) &= \sum_x (k^* P_1(x) + (1 - k^*) P_2(x)) \\ &= k^* \sum_x P_1(x) + (1 - k^*) \sum_x P_2(x) \\ &= k^* + 1 - k^* = 1 \end{aligned}$$

- Folgt direkt aus der Eigenschaft der Linearität des Erwartungswertes.
- Sei m_k das k -te Moment der Mischverteilung von X , dann gilt $m_2 = E(X)^2$ und somit $m_2 = \text{Var}(X) + (E(X))^2 = \sigma^2 + (m_1)^2$. Es folgt:

$$\begin{aligned} \sigma^2 &= m_2 - (m_1)^2 \\ &= k^* (\sigma_1^2 + \mu_1^2) + (1 - k^*) (\sigma_2^2 + \mu_2^2) - (k^* \mu_1 + (1 - k^*) \mu_2)^2 \\ &= k^* \sigma_1^2 + (1 - k^*) \sigma_2^2 + k^* \mu_1^2 + (1 - k^*) \mu_2^2 - (k^{*2} \mu_1^2 + (1 - k^*)^2 \mu_2^2 + 2k^* (1 - k^*) \mu_1 \mu_2) \\ &= k^* \sigma_1^2 + (1 - k^*) \sigma_2^2 + k^* (1 - k^*) (\mu_1 - \mu_2)^2 \end{aligned}$$

□

Sei $P(x; \theta) = k^* P_1(x; \theta_1) + (1 - k^*) P_2(x; \theta_2)$ die Wahrscheinlichkeitsfunktion einer diskreten Mischverteilung einer einparametrischen Zufallsvariablen mit Parameter θ_i , $i = 1, 2$, und seien im Folgenden $x = (x_1, \dots, x_n)$ die Beobachtungen, an die die Mischverteilung angepasst werden soll, dann ist die Likelihood der Mischverteilung durch folgenden Ausdruck gegeben:

$$L(\theta_1, \theta_2 | x, k^*) = \prod_{i=1}^n [k^* P_1(x_i; \theta_1) + (1 - k^*) P_2(x_i; \theta_2)] \quad (3.3)$$

Schätzer für die Parameter θ_i werden durch gewohnte Verfahren ermittelt. Mit der vorgestellten Definition der Mischverteilung und deren Eigenschaften wird nun ein Verfahren zur verblindeten Fallzahlanpassung hergeleitet.

Die grundlegende Idee der verblindeten Fallzahlanpassung im Zweistichprobenfall ist die Betrachtung einer aus den einzelnen Wahrscheinlichkeitsfunktionen entstehenden Mischverteilung. Durch die benötigte Annahme an die Effektgröße wird die zu schätzende Parameteranzahl in der vorliegenden Problemstellung reduziert. Im Folgenden wird die Idee zur verblindeten Fallzahladjustierung an Poisson-verteilten Daten vorgestellt.

Seien die Zufallsvariablen $N_{ij} \sim \text{Pois}(t_{ij}\lambda_i)$, $i = 0, 1$, verteilt. Im Folgenden wird davon ausgegangen, dass der Randomisierungsfaktor k bekannt ist. Die Mischwahrscheinlichkeitsfunktion hat die Form

$$p(n_{ij}) = \frac{1}{k+1} p_{(\lambda_0)}(n_{ij}) + \frac{k}{k+1} p_{(\lambda_0 \exp(\beta_1))}(n_{ij}).$$

Der Erwartungswert der Mischverteilung ist durch den Ausdruck $\lambda := \frac{\lambda_0}{k+1} (1 + k \exp(\beta_1))$ gegeben (Satz 3.1). Wird die Kontrollrate aus den verblindeten Daten geschätzt, was eine Zuordnung der Beobachtungen in Behandlungs oder Kontrollgruppe nicht möglich macht, so kann die Mischverteilung zur Schätzung der Parameter herangezogen werden. Die Likelihood für die Mischverteilung lässt sich mittels Gl. 3.3 zu

$$\begin{aligned} L(\lambda_0, \beta_1 | k, n_{ij}) &= \prod_{i=0}^1 \prod_{j=1}^{m_i} \left(\frac{1}{k+1} (n_{ij}!)^{-1} (t_{ij}\lambda_0)^{n_{ij}} \exp(-t_{ij}\lambda_0) \right. \\ &\quad \left. + \frac{k}{k+1} (n_{ij}!)^{-1} (t_{ij}\lambda_0 \exp(\beta_1))^{n_{ij}} \exp(-t_{ij}\lambda_0 \exp(\beta_1)) \right) \end{aligned} \quad (3.4)$$

ableiten. Da die Effektgröße β_1 in der verblindeten Analyse nicht bekannt ist, wird eine Annahme an diese Größe benötigt. Infolgedessen wird der wahre Effekt durch $\beta_1^* = \beta_1$ in der Likelihood in Gl. 3.4 ersetzt. Die Log-Likelihood hat damit die folgende Form:

$$\begin{aligned} l(\lambda_{0(b)}, |k, \beta_1^*, n_{ij}) &= \sum_{i=0}^1 \sum_{j=1}^{m_i} \log \left(\frac{1}{k+1} \right) + \log((n_{ij}!)^{-1}) + n_{ij} \log(t_{ij}) + n_{ij} \log(\lambda_{0(b)}) \\ &\quad + \log \left([\exp(-t_{ij}\lambda_{0(b)}) + k \exp(\beta_1^*)^{n_{ij}} \exp(-t_{ij}\lambda_{0(b)} \exp(\beta_1^*))] \right) \end{aligned}$$

Da keine analytische Lösung, die zu einer Maximierung der Log-Likelihood führt, existiert, wird auf numerische Verfahren zur Optimierung zurückgegriffen. Der Schätzer $\lambda_{0(b)}$ besitzt für balancierte Nachbeobachtungszeiten die gewohnten Maximum-Likelihood Eigenschaften und wird im Fallzahlanpassungsschritt in die approximative Fallzahlformel für die Kontrollgruppe in Gl. 2.14 eingesetzt, was zu einer adjustierten Fallzahl führt.

Da dieses Verfahren nur mittels numerischer Optimierung durchzuführen ist, wird im weiteren Verlauf ein anderer Satz von Friede und Schmidli (2010a) vorgestellt. Die Kenntnis des Randomisierungsfaktors und die Spezifizierung der Effektgröße werden ebenfalls als Voraussetzungen angesehen. Wie erwähnt ist der Erwartungswert der Mischverteilung durch $\lambda := \lambda_0 / (k+1) (1 + \exp(\beta_1))$ gegeben. Um die Gesamtrate λ zu schätzen, verwenden Friede und Schmidli (2010a) die Momenten-

methode, wobei eine anschließende Adjustierung mit dem reziproken Wert von $(1 + \exp(\beta_1^*)) / (k + 1)$ zu einem verblindeten Schätzer für λ_0 führt. Im Adjustierungsschritt ist zu beachten, dass der angenommene Effekt β_1^* anstatt dem wahren β_1 verwendet wird. Ein verblindeter Schätzer für die Kontrollrate ist demzufolge durch

$$\hat{\lambda}_{0(b)} = \frac{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} n_{ij}}{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} t_{ij}} \left(\frac{k+1}{k \exp \beta_1^* + 1} \right) \quad (3.5)$$

mit einem angenommenen Effekt von β_1^* gegeben. Im Folgenden werden für diesen Schätzer asymptotische Eigenschaften nachgewiesen.

Satz 3.2 *Seien $\beta_1 = \beta_1^*$ und die Summe der Nachbeobachtungszeiten balanciert (in Formeln $T_0^* := \sum_{j=1}^{m_0^*} t_{0j} = \sum_{j=1}^{km_0^*} t_{1j} = kT_1^*$), so ist der Schätzer aus Gl. 3.5 unter der Voraussetzung 2.1 erwartungstreu und L_2 konsistent.*

BEWEIS. Zum Beweis der Erwartungstreu:

$$\begin{aligned} E(\hat{\lambda}_{0(b)}) &= E \left(\frac{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} N_{ij}}{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} t_{ij}} \left(\frac{k+1}{k \exp(\beta_1^*) + 1} \right) \right) \\ &= \frac{1}{t_0^* + k t_1^*} \left(\frac{k+1}{1 + k \exp(\beta_1^*)} \right) E \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} N_{ij} \right) \\ &= \frac{1}{t_0^* (1+k)} \left(\frac{k+1}{1 + k \exp(\beta_1^*)} \right) (T_0^* \lambda_0 + T_1^* \lambda_1) \\ &= \frac{1}{1+k} \left(\frac{k+1}{1 + k \exp(\beta_1^*)} \right) \lambda_0 (1 + k \exp(\beta_1^*)) \\ &= \lambda_0 \end{aligned}$$

Die Konsistenz des Schätzers wird durch den mittleren quadratischen Fehler (MSE) gezeigt. Da für $\beta_1^* = \beta_1$ der Schätzer erwartungstreu ist, hat der MSE des Schätzers folgende Darstellung:

$$\begin{aligned} \text{MSE}(\hat{\lambda}_{0(b)}) &= \text{Var}(\hat{\lambda}_{0(b)}) \\ &= \frac{1}{(T_0^* + T_1^*)^2} \left(\frac{k+1}{1 + k \exp(\beta_1^*)} \right)^2 \text{Var} \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} N_{ij} \right) \\ &= \frac{1}{(T_0^* + T_1^*)^2} \left(\frac{k+1}{1 + k \exp(\beta_1^*)} \right)^2 (T_0^* \lambda_0 + T_1^* \lambda_1) \\ &= \frac{1}{T_0^{2*}} \left(\frac{1}{1 + k \exp(\beta_1^*)} \right)^2 T_0^* \lambda_0 (1 + k \exp(\beta_1^*)) \\ &= \frac{1}{T_0^*} \left(\frac{1}{1 + k \exp(\beta_1^*)} \right) \lambda_0 \end{aligned}$$

Unter der Voraussetzung 2.1 folgt $\text{MSE}(\hat{\lambda}_{0(b)}) \rightarrow 0$. □

Die Betrachtung einer Mischverteilung ist für theoretische Überlegungen nützlich, weist jedoch in der praktischen Anwendung negative Aspekte auf, die bereits erläutert wurden. Somit wird in den

folgenden Simulationsstudien der Fokus auf den vorgestellten Ansatz von Friede und Schmidli (2010a) gelegt.

3.2.2 Unverblindeter Ansatz zur Fallzahlenanpassung

Ein unverblindetes Verfahren zur Fallzahlenanpassung im Poisson-Modell wird von Schneider et al. (2013a) gegeben. Die Kontrollrate wird unmittelbar aus den Beobachtungen der Kontrollgruppe der internen Pilotstudie geschätzt. Der Momentenschätzer ist durch

$$\hat{\lambda}_{0(u)} = \frac{\sum_{j=1}^{m_0^*} n_{0j}}{\sum_{j=1}^{m_0^*} t_{0j}} \quad (3.6)$$

gegeben. Folgender Satz sichert die Erwartungstreue und Konsistenz.

Satz 3.3 *Der Schätzer aus Gl. 3.6 ist mit der Voraussetzung 2.1 erwartungstreu und L_2 konsistent.*

BEWEIS. Zum Beweis der Erwartungstreue:

$$\begin{aligned} E(\hat{\lambda}_{0(u)}) &= E\left(\frac{\sum_{j=1}^{m_0^*} N_{0j}}{\sum_{j=1}^{m_0^*} t_{0j}}\right) \\ &= \frac{1}{\sum_{j=1}^{m_0^*} t_{0j}} E\left(\sum_{j=1}^{m_0^*} N_{0j}\right) \\ &= \frac{1}{\sum_{j=1}^{m_0^*} t_{0j}} \sum_{j=1}^{m_0^*} t_{0j} \lambda_0 \\ &= \lambda_0 \end{aligned}$$

Um die Konsistenz des Schätzers zeigen zu können, wird der mittlere quadratische Fehler (MSE) herangezogen. Der MSE des Schätzers ist durch

$$\text{MSE}(\hat{\lambda}_{(u)}) = \frac{\lambda_0}{t_0^*}$$

gegeben. Unter der Voraussetzung 2.1 konvergiert $\text{MSE}(\hat{\lambda}_{(u)}) \rightarrow 0$. □

Der verblindete bzw. entblindete Schätzer aus Gl. 3.5 und Gl. 3.6 wird im Fallzahlenanpassungsschritt in die approximative Fallzahlformel für die Kontrollgruppe in Gl. 2.14 eingesetzt, was zu einer adjustierten Fallzahl $\hat{m}_{0(u)}$ bzw. $\hat{m}_{0(b)}$ führt.

3.3 Fallzahladjustierung im Negativ-Binomial-Modell

3.3.1 Verblindete Ansätze zur Fallzahlenanpassung

Dem Folgenden liegt das Negativ-Binomial-Modell aus Gl. 2.7 zugrunde, für das Ansätze zur Fallzahladjustierung vorgestellt werden. Notationen und Ideen können aus Abschnitt 3.2 übernommen

werden. Vergleichbar zum vorherigen Abschnitt wird zunächst ein Zugang zur verblindeten Fallzahl-anpassung mittels diskreter Mischung von zwei Negativ-Binomial-Verteilungen gesucht. Seien die diskreten Zufallsvariablen $N_{ij} \sim \text{NB}(t_{ij}\lambda_0 \exp(\beta_1 x_i), \varphi)$, $i = 0, 1$, $j = 1, \dots, m_i$, verteilt. Es wird davon ausgegangen, dass der Randomisierungsfaktor k bekannt ist. Die Mischwahrscheinlichkeitsfunktion der einzelnen Wahrscheinlichkeitsfunktionen $p(\cdot)$ hat die Form

$$p(n_{ij}) = \frac{1}{k+1} p_{(\lambda_0, \varphi)}(n_{ij}) + \frac{k}{k+1} p_{(\lambda_0 \exp(\beta_1), \varphi)}(n_{ij}).$$

Der Erwartungswert lässt sich auch im Negativ-Binomial-Modell als Summe der gewichteten Erwartungswerte schreiben und wird zu $\lambda := \lambda_0 / (k+1) (1 + k \exp(\beta_1))$ berechnet (Satz 3.1).

Das Interesse liegt in den verblindeten Schätzern der Parameter λ_0 und φ , die durch Maximierung der folgenden Likelihood ermittelt werden:

$$\begin{aligned} L(\lambda_0, \varphi | k, \beta_1^*, n_{ij}) = & \prod_{i=0}^1 \prod_{j=1}^{m_i} \left[\left(\frac{1}{k+1} \right) \frac{\Gamma(n_{ij} + \varphi^{-1})}{n_{ij}! \Gamma(\varphi^{-1})} \left(\frac{\varphi t_{ij} \lambda_0}{1 + \varphi t_{ij} \lambda_0} \right)^{n_{ij}} \left(\frac{1}{1 + \varphi t_{ij} \lambda_0} \right)^{\varphi^{-1}} \right. \\ & \left. + \left(\frac{k}{k+1} \right) \frac{\Gamma(n_{ij} + \varphi^{-1})}{n_{ij}! \Gamma(\varphi^{-1})} \left(\frac{\varphi t_{ij} \lambda_0 \exp(\beta_1^*)}{1 + \varphi t_{ij} \lambda_0 \exp(\beta_1^*)} \right)^{n_{ij}} \left(\frac{1}{1 + \varphi t_{ij} \lambda_0 \exp(\beta_1^*)} \right)^{\varphi^{-1}} \right]. \end{aligned} \quad (3.7)$$

Es ist zu beachten, dass der wahre Effekt β_1 durch den angenommenen Effekt β_1^* ersetzt wurde. Die Schätzer $\hat{\lambda}_0$ und $\hat{\varphi}$ ergeben sich durch Lösung des Gleichungssystems $\left(\frac{\partial L(\lambda_0, \varphi | k, \beta_1^*, n_{ij})}{\partial \lambda_0}, \frac{\partial L(\lambda_0, \varphi | k, \beta_1^*, n_{ij})}{\partial \varphi} \right)' = \mathbf{0}$. Gilt $\beta_1 = \beta_1^*$ und sind die Nachbeobachtungszeiten balanciert, so sind Schätzer, die aus der Nullsetzung der Log-Likelihood numerisch bestimmt werden, asymptotisch erwartungstreu.

Der Fokus in dieser Arbeit wird jedoch auf einen anderen Ansatz zur Fallzahlanpassung, der durch Friede und Schmidli (2010b) vorgestellt wird, gelegt. Vergleichbar zum Abschnitt 3.2.1 wird ein verblindeter Schätzer für die Gesamtrate mittels der Maximierung der Log-Likelihood in Gl. 2.8 erhalten, wobei der Behandlungsindikator x_i vernachlässigt wird (d.h. $x_i = 0, \forall i$). Die zugehörige Likelihood hat folgende Darstellung

$$L(\lambda, \varphi | n_{ij}, t_{ij}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} \frac{\Gamma(n_{ij} + \varphi^{-1})}{n_{ij}! \Gamma(\varphi^{-1})} \left(\frac{\varphi t_{ij} \lambda}{1 + \varphi t_{ij} \lambda} \right)^{n_{ij}} \left(\frac{1}{1 + \varphi t_{ij} \lambda} \right)^{\varphi^{-1}},$$

ist jedoch per Definition nicht die zugehörige Likelihood der Mischverteilung der einzelnen Wahrscheinlichkeitsfunktionen (vgl. Gl. 3.7).

Die Schätzer ergeben sich abermals durch die numerische Lösung des Gleichungssystems $\left(\frac{\partial L(\lambda, \varphi | k, \beta_1^*, n_{ij})}{\partial \lambda}, \frac{\partial L(\lambda, \varphi | k, \beta_1^*, n_{ij})}{\partial \varphi} \right)' = \mathbf{0}$. Die Eigenschaften der Schätzer resultieren aus der Maximum-Likelihood Methode und beruhen auf der Annahme über den zugrunde liegenden Prozess. Das bedeutet, dass die Verteilung bzw. Dichtefunktion der Zufallsvariable richtig spezifiziert werden muss. Ist die Verteilungsannahme inkorrekt, so können inkonsistente Schätzer erhalten werden. Dieser Fall tritt durch das Ignorieren der Gruppenzugehörigkeit auf, so dass die vertrauten Eigenschaften der asymptotischen Erwartungstreu des Shape-Parameter-Schätzers nicht gewährleistet sind. Die resultierende Verzerrung steigt bei wachsendem Effekt β_1 . Lediglich unter der Nullhypothese ist der

Schätzer asymptotisch erwartungstreu. Für den verblindeten Schätzer der Kontrollrate kann dieses Verhalten nicht beobachtet werden. Nach Adjustierung des Gesamtratenschätzers $\hat{\lambda}$ mit dem Faktor $\frac{k+1}{k \exp(\beta_1^*)+1}$ wird für balancierte Nachbeobachtungszeiten $T_0^* = kT_1^*$ ein asymptotisch erwartungstreuer Schätzer $\hat{\lambda}_{0(b)} := \hat{\lambda} \frac{k+1}{k \exp(\beta_1^*)+1}$ erhalten. Für die starke Annahme $t_{ij} = t$ führt die Maximierung der Log-Likelihood zu einer geschlossenen Lösung. Der verblindete Schätzer hat die folgende Form:

$$\begin{aligned}\hat{\lambda}_{0(b)} &= \frac{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} n_{ij}}{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} t} \left(\frac{k+1}{k \exp(\beta_1^*)+1} \right) \\ &= \frac{\sum_{i=0}^1 \sum_{j=1}^{m_i^*} n_{ij}}{(m_0^* + m_1^*) t} \left(\frac{k+1}{k \exp(\beta_1^*)+1} \right).\end{aligned}$$

Der Beweis zur Erwartungstreue und Konsistenz ist hierbei analog zu Satz 3.2. Im Fallzahlenanpassungsschritt werden die verblindeten Schätzer $\hat{\lambda}_{0(b)}$ und $\hat{\phi}_{0(b)}$ in die approximative Fallzahlformel aus Gleichung 2.15 eingesetzt.

3.3.2 Unverblindeter Ansatz zur Fallzahlenanpassung

Ein unverblindetes Verfahren zur Fallzahlenanpassung im Negativ-Binomial-Modell wird durch Cook et al. (2009) beschrieben. Unverblindete Schätzer für die Kontrollrate und den Shape-Parameter werden durch die Maximierung der folgenden Einstichproben-Likelihood erhalten.

$$\begin{aligned}l(\lambda_0, \varphi | n_{0j}, t_{0j}) &= \sum_{j=1}^{m_0} \log(\Gamma((n_{0j} + \varphi^{-1})) - \log(\Gamma(\varphi^{-1})) \\ &\quad + n_{0j} [\log(\varphi t_{0j} \lambda_0) - \log(1 + \varphi t_{0j} \lambda_0)] - \varphi^{-1} \log(1 + \varphi t_{0j} \lambda_0)\end{aligned}\tag{3.8}$$

Die Ableitungen der Log-Likelihood in Gl. 3.8 sind durch die Gleichungen

$$\begin{aligned}\frac{\partial l(\lambda_0, \varphi | n_{0j}, t_{0j})}{\partial \lambda_0} &= \sum_{j=1}^{m_0} \frac{n_{0j} - t_{0j} \lambda_0}{\lambda_0 (1 + \varphi t_{0j} \lambda_0)} \\ \frac{\partial l(\lambda_0, \varphi | n_{0j}, t_{0j})}{\partial \varphi} &= \sum_{j=1}^{m_0} \left[\left(\sum_{i=0}^{n_{0j}^*} \frac{i}{1 + \varphi i} \right) + \varphi^{-2} \log(1 + \varphi t_{0j} \lambda_0) - \frac{(n_{0j} + \varphi^{-1}) t_{0j} \lambda_0}{1 + \varphi t_{0j} \lambda_0} \right]\end{aligned}$$

gegeben. Der Schätzer für die Kontrollrate und den Shape-Parameter sind asymptotisch erwartungstreu und konsistent (vgl. Satz 2.3 und Satz 2.4). Die Einführung eines Gruppenindicators führt zu der Log-Likelihood Darstellung, die in Gleichung 2.8 gegeben ist, und liefert einen alternativen Schätzer für den Shape-Parameter, der zur Anpassung in der Zwischenstufe benutzt wird. Im Folgenden wird der aus dem Zweistichprobenmodell resultierende Schätzer angesichts einer geringeren Variabilität zur Fallzahlplanung herangezogen. Für die Schätzer gelten die Eigenschaften, die in Satz 2.4 dargestellt sind. Die unverblindeten Schätzer $\hat{\lambda}_{0(u)}$ und $\hat{\phi}_{0(u)}$ werden im Anpassungsschritt in die approximative Fallzahlformel aus Gleichung 2.15 eingesetzt und führen zu einer adjustierten Fallzahl $\hat{m}_{0(u)}$.

3.4 Fallzahladjustierung für den Quasi-Likelihood Ansatz

3.4.1 Verblindeter Ansatz zur Fallzahlanpassung

Da der Schätzer für die Kontrollrate identisch zu dem Schätzer aus dem Poisson-Regressionmodell ist, wird der verblindete Schätzer der Kontrollrate λ_0 analog zu Abschnitt 3.2.1 berechnet (Friede und Schmidli, 2010a). Ein verblindeter Momentenschätzer für σ^2 wird für gleiche Nachbeobachtungszeiten von Friede und Schmidli (2010a) gegeben. Ein verallgemeinerter Schätzer ist im folgenden Ausdruck gegeben:

$$\hat{\sigma}_{(b)}^2 = \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} \left(\frac{T}{T t_{ij} - t_{ij}^2} \right) \frac{(n_{ij} - t_{ij} \hat{\lambda}_{(b)})^2}{\hat{\lambda}_{(b)}} \right) / (m_0^* + m_1^*)$$

mit $\hat{\lambda}_{(b)} = \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} n_{ij} \right) / \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} t_{ij} \right)$ und $T = \sum_{i=0}^1 \sum_{j=1}^{m_i^*} t_{ij}$. Der Zähler des Ausdrucks $\sum_{i=0}^1 \sum_{j=1}^{m_i^* + m_1^*} (n_{ij} - \hat{\lambda})^2 / (m_0^* + m_1^* - 1)$ schätzt erwartungstreu die Gesamtstreuung, die sich jedoch aus der Varianz innerhalb der Gruppen sowie zwischen den Gruppen zusammensetzt. Demzufolge wird σ^2 von dem gegebenen Schätzer $\hat{\sigma}_{(b)}^2$ in Abhängigkeit von der Effektgröße β_1 überschätzt und ist lediglich unter Hypothese $\beta_1 = 0$ asymptotisch erwartungstreu. Dieses Phänomen ist ebenso bei der Fallzahlanpassung bei normalverteilten Daten zu beobachten. Weitere Untersuchungen zeigen jedoch, dass die Überschätzung der Varianz in klinisch relevanten Situationen zu keinem Nachteil führt und die gewünschten Eigenschaften der Prozeduren hinsichtlich der Power und Fehler 1. Art erreicht bzw. kontrolliert werden (Friede und Schmidli, 2010a; Friede und Kieser, 2013; Schneider et al., 2013c,a). Die verblindeten Schätzer $\hat{\lambda}_{0(b)}$ und $\hat{\sigma}_{(b)}^2$ werden im Zwischenschritt in die approximative Fallzahlformel aus Gleichung 2.16 gesetzt und führen zu einer adjustierten Fallzahl $\hat{m}_{0(b)}^{QL}$.

3.4.2 Unverblindeter Ansatz zur Fallzahlanpassung

Im unverblindeten Ansatz wird die Varianz σ^2 aus beiden Behandlungsgruppen geschätzt. Der Schätzer ist durch

$$\hat{\sigma}_{(u)}^2 = \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^* + m_1^*} \left(\frac{T_i}{T_i t_{ij} - t_{ij}^2} \right) \frac{(n_{ij} - t_{ij} \hat{\lambda}_{0(u)})^2}{\hat{\lambda}_{0(u)}} \right) / (m_0^* + m_1^*)$$

gegeben und wird im Anpassungsschritt mit dem Schätzer $\hat{\lambda}_{0(u)}$ für die Kontrollrate in Gl. 3.6 in die approximative Fallzahlformel in Gl. 2.16 eingesetzt und führt zu einer adjustierten Fallzahl $\hat{m}_{0(u)}^{QL}$.

3.5 EM-Algorithmus-basierender Ansatz zur verblindeten Fallzahlanpassung

3.5.1 Einführung

Der EM-Algorithmus ist ein iteratives Verfahren zur Maximum-Likelihood-Schätzung bei unvollständigen Daten, welches erstmals von Dempster et al. (1977) eingeführt wurde. Die Unvollständigkeit der Daten kann durch nicht beobachtbare Werte erklärt werden und tritt häufig bei der

Betrachtung von Mischverteilungen in den Vordergrund. Letztere sind keine fehlenden Werte per Definition, sondern folgen aus der Mischungsoperation mit einem nicht beobachtbaren zufälligen Effekt. In diesem Zusammenhang wird eine ausführliche Beschreibung der Prozedur durch Dempster et al. (1977) und explizit für Poisson-Mischverteilungen durch Karlis (2001) bereitgestellt.

Wird der EM-Algorithmus in der verblindeten Fallzahlanpassung angewendet, so besteht die weitere Unvollständigkeit der Daten darin, dass die Gruppenzugehörigkeit der in der Studie eingeschlossenen Patienten nicht bekannt ist. Ein Fallzahlanpassungsverfahren wurde für den Fall der Mischung von zwei Normalverteilungen nach Gould und Shih (1992) beschrieben. Eine detaillierte Untersuchung der Prozedur und ihrer Eigenschaften wies grundlegende Defizite in praktisch relevanten Szenarien auf (Friede und Kieser, 2002; Waksman, 2007). Trotz der Mängel wird in diesem Abschnitt eine EM-Algorithmus basierte Prozedur nach Cook et al. (2009) zur verblindeten Störparameterschätzung im Poisson-Gamma Mischverteilungsmodell vorgestellt und analysiert. Bevor die Merkmale des EM-Algorithmus-basierenden Ansatzes für Zähldaten überprüft werden, wird auf die einzelnen Operationsschritte eingegangen. Herleitungen und Beweise für die Funktionsweise des EM-Algorithmus werden in dieser Arbeit nicht gegeben. Es wird auf Dempster et al. (1977) verwiesen.

3.5.2 EM-Algorithmus für das Poisson-Gamma-Regressionsmodell

Seien die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, mit individueller Rate $u_{ij}\lambda_i$ Poisson-verteilt, wobei u_{ij} eine Realisierung der Zufallsvariable U_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, darstellt, die einer Gamma-Verteilung mit Erwartungswert 1 und Varianz φ folgt. Es sei $X = (X_{0j}, \dots, X_{1m_0+m_1})^T$ der Vektor der Behandlungszugehörigkeit, wobei $X_{ij} = 1$, (0) die Zugehörigkeit in die Behandlungsgruppe (Kontrollgruppe) bezeichnet. Die Zufallsvariable X_{ij} sei Bernoulli-verteilt mit Parameter q , d.h. $P(X_i = 1) = q$. Dann folgt für die Likelihood (Cook et al., 2009):

$$L(\lambda_0, \beta_1, \varphi | u_{ij}, t_{ij}, x_i, n_{ij}) = \prod_{i=0}^1 \prod_{j=1}^{m_{ij}} \frac{(t_{ij} u_{ij} \lambda_0 \exp(\beta_1 x_i))^{n_{ij}} \exp(-t_{ij} u_{ij} \lambda_0 \exp(\beta_1 x_i))}{n_{ij}!} \cdot \frac{u_{ij}^{\varphi^{-1}-1} \exp(-u_{ij} \varphi^{-1})}{\Gamma(\varphi^{-1}) \varphi^{\varphi^{-1}}} q^{x_i} (1-q)^{1-x_i}.$$

Sei im Folgenden $L(\lambda_0, \beta_1, \varphi | u_{ij}, t_{ij}, x_i, n_{ij}) = L(\theta | u_{ij}, t_{ij}, x_i, n_{ij})$ und $l(\theta) := \log(L(\theta | u_{ij}, t_{ij}, x_i, n_{ij}))$. Dann ist die Log-Likelihood durch

$$l(\theta) = \sum_{i=0}^1 \sum_{j=1}^{m_{ij}} [n_{ij}(\log(u_{ij}) + \log(t_{ij})) + n_{ij}(\log(\lambda_0) + \beta_1 x_i) - (t_{ij} u_{ij} \lambda_0 \exp(\beta_1 x_i)) - \log(n_{ij}!)] \\ + (\varphi^{-1} - 1)\log(u_{ij}) - u_{ij} \varphi^{-1} - \log(\Gamma(\varphi^{-1})) - (\varphi^{-1})\log(\varphi) \\ + x_i \log(q) + (1 - x_i) \log(1 - q)]$$

gegeben. Die Log-Likelihood wird in zwei Teile gegliedert, wobei l_1 von λ_0 und β_1 und l_2 von φ abhängen und für die Maximierung unwichtiger Komponenten nicht weiter aufgeführt werden.

$$l_1(\theta) = \sum_{i=0}^1 \sum_{j=1}^{m_{ij}} [n_{ij}(\log \lambda_0 + \beta_1 x_i) - (t_{ij} u_{ij} \lambda_0 \exp(\beta_1 x_i))] \quad (3.9)$$

$$l_2(\theta) = \sum_{i=0}^1 \sum_{j=1}^{m_{ij}} [(\varphi^{-1} - 1) \log(u_{ij}) - u_{ij} \varphi^{-1} - \log(\Gamma(\varphi^{-1})) - (\varphi^{-1}) \log(\varphi)] \quad (3.10)$$

Der EM-Algorithmus ist ein iterativer Algorithmus, der aus zwei sich abwechselnden Schritten, dem E- und M-Schritt, besteht. Im sogenannten Expectation-Schritt werden die Erwartungswerte der unbekannten Variablen bedingt auf die Daten und die Parameterschätzer aus der vorhergehenden Iteration berechnet. Im Maximization-Schritt werden die bedingten Erwartungswerte genutzt, um Maximum-Likelihood Schätzer der gewünschten Parameter für den vollständigen Datensatz zu berechnen. Formal kann der EM-Algorithmus für den $k + 1$ -ten Iterationsschritt wie folgt beschrieben werden:

- **E-Schritt:** Sei $Q(\theta|\theta_{(k)})$ die bedingte Erwartung der Log-Likelihood in der $k + 1$ -ten Iteration des Expectation-Schritts ($Q(\theta|\theta_{(k)}) = E(l(\theta)|\theta_{(k)})$). Die bedingte Erwartung kann als Summe der Log-Likelihood Funktionen aus Gl. 3.9 und Gl. 3.10 zu $Q(\theta|\theta_{(k)}) = Q_1(\theta; \hat{\theta}_{(k)}) + Q_2(\theta; \hat{\theta}_{(k)})$ berechnet werden. Die bedingten Erwartungswerte sind durch

$$\begin{aligned} Q_1(\theta|\hat{\theta}_{(k)}) &= E(l_1(\lambda_0, \beta_1)|\theta_{(k)}) \\ &= \sum_{i=0}^1 \sum_{j=1}^{m_{ij}} [n_{ij}(\log(\lambda_0) + \beta_1 \zeta_{ij(k)}) - (1 - \zeta_{ij(k)}) \eta_{0j(k)} \lambda_0 t_{ij} - \zeta_{ij(k)} \eta_{1j(k)} \lambda_0 \exp(\beta_1) t_{ij}] \end{aligned}$$

und

$$\begin{aligned} Q_2(\theta|\hat{\theta}_{(k)}) &= E(l_2(\varphi)|\theta_{(k)}) \\ &= \sum_{i=0}^1 \sum_{j=1}^{m_{ij}} [(\varphi^{-1} - 1)(\gamma_{0j(k)}(1 - \zeta_{ij(k)}) + \gamma_{1j(k)} \zeta_{ij(k)}) - (\eta_{0j(k)}(1 - \zeta_{ij(k)}) + \eta_{1j(k)} \zeta_{ij(k)}) \varphi^{-1} \\ &\quad - \log(\Gamma(\varphi^{-1})) - (\varphi^{-1}) \log(\varphi)] \end{aligned}$$

gegeben. In den Ausdrücken Q_1 und Q_2 werden die nicht beobachtbaren Größen u_{ij} und

$\log(u_{ij})$ durch die folgenden bedingten Erwartungswerte ersetzt:

$$\eta_{0j(k)} = E(u_{ij}|X_i = 0, n_{ij}, \hat{\theta}_{(k)}) = \frac{1 + \varphi_{(k)} n_{ij}}{1 + \varphi_{(k)} \lambda_{0(k)} t_{ij}} \quad (3.11)$$

$$\eta_{1j(k)} = E(u_{ij}|X_i = 1, n_{ij}, \hat{\theta}_{(k)}) = \frac{1 + \varphi_{(k)} n_{ij}}{1 + \varphi_{(k)} \lambda_{0(k)} \exp(\beta_{1(k)} x_i) t_{ij}} \quad (3.12)$$

$$\gamma_{0j(k)} = E(\log(u_{ij})|X_i = 0, n_{ij}, x_i, \hat{\theta}_{(k)}) = \Psi(n_{ij} + \varphi_{(k)}^{-1}) + \log\left(\frac{\varphi_{(k)}}{1 + \varphi_{(k)} \lambda_{0(k)} t_{ij}}\right) \quad (3.13)$$

$$\begin{aligned} \gamma_{1j(k)} &= E(\log(u_{ij})|X_i = 1, n_{ij}, x_i, \hat{\theta}_{(k)}) \\ &= \Psi(n_{ij} + \varphi_{(k)}^{-1}) + \log\left(\frac{\varphi_{(k)}}{1 + \varphi_{(k)} \lambda_{0(k)} \exp(\beta_{1(k)} x_i) t_{ij}}\right) \end{aligned} \quad (3.14)$$

Der bedingte Erwartungswert der Gruppenzugehörigkeit dient als Gewichtungsfaktor und hat die Form

$$\begin{aligned} \zeta_{ij(k)} &= P(X_{ij} = 1|n_{ij}, \hat{\theta}_{(k)}) \\ &= \left[1 + \frac{1-q}{q} \left(\frac{1}{\exp(\beta_{1(k)})} \right)^{n_{ij}} \left(\frac{1 + \varphi_{(k)} \lambda_{0(k)} t_{ij}}{1 + \varphi_{(k)} \lambda_{0(k)} + \exp(\beta_{1(k)}) t_{ij}} \right)^{-(n_{ij} + \varphi_{(k)}^{-1})} \right]^{-1}. \end{aligned} \quad (3.15)$$

Der Beweis der bedingten Erwartungswerte (Gl. 3.11-Gl. 3.14) folgt mit dem Lemma von Sapatinas (1995), das in Karlis (2001) gegeben ist. Gleichung 3.15 lässt sich mit Hilfe des Bayes-Theorems verifizieren (Anhang A.3).

- **M-Schritt:** Die vollständigen erwarteten Log-Likelihood Funktionen $Q_1(\theta; \hat{\theta}_{(k)})$ und $Q_2(\theta; \hat{\theta}_{(k)})$ können im $k+1$ -ten Iterationsschritt separat maximiert werden. Numerische Optimierungsverfahren wie der Newton-Raphson Algorithmus oder Nelder-Mead (Himmelblau, 1972, Kapitel 4) können für die Maximierung von $Q_2(\theta; \hat{\theta}_{(k)})$ und für die mehrdimensionale Funktion $Q_1(\theta; \hat{\theta}_{(k)})$ verwendet werden. Eine weitere Möglichkeit, $Q_1(\theta; \hat{\theta}_{(k)})$ zu maximieren, wird von Cook et al. (2009) beschrieben. Als Resultat wird ein neuer Parametervektor $\hat{\theta}_{(k+1)} = (\hat{\lambda}_{0(k+1)}, \hat{\beta}_{1(k+1)}, \hat{\varphi}_{(k+1)})$ erhalten, der erneut zur Berechnung der Erwartungswerte im E-Schritt benutzt wird.

Der Expectation und Maximization-Schritt wird wiederholt, bis ein gegebenes Konvergenzkriterium erfüllt ist. In einer verblindeten Fallzahlanpassungsprozedur werden die erhaltenen Schätzer der Kontrollrate λ_0 und des Shape-Parameters φ in die approximative Fallzahlformel in Gl. 2.15 eingesetzt und führen zu einer erneuten Fallzahl $\hat{n}_{0(b)}$.

Im Folgenden wird auf die Wahl des Konvergenzkriteriums und die Spezifizierung der Startwerte eingegangen, die einen wesentlichen Einfluss auf die Anzahl der Iterationen des EM-Algorithmus und die resultierenden Schätzer haben.

Konvergenzkriterium: Der EM-Algorithmus konvergiert oftmals recht langsam. Ein zu frühzeitiges

Beenden der Prozedur kann zu verzerrten Schätzern führen. Aus diesem Grund sollte ein geeignetes Konvergenzkriterium gewählt werden (Dempster et al., 1977; McLachlan und Krishnan, 1997; Karlis, 2001). Es gibt verschiedene Möglichkeiten ein Konvergenzkriterium zu definieren. In dieser Arbeit wird eine Bedingung an die Differenz aufeinanderfolgender Parameterschätzer gemacht:

$$\max_{1 \leq i \leq 3} \{|\hat{\theta}_{i(k+1)} - \hat{\theta}_{i(k)}|\} < \delta$$

für einen vordefinierten Wert δ .

Startwerte: Des Weiteren ist bekannt, dass die resultierenden Parameterschätzer von der Wahl der Startwerte abhängen. Im Kontext der Fallzahlanpassung ist diese Eigenschaft für normalverteilte Daten von Friede und Kieser (2002) beschrieben. Die für die vorliegende Arbeit festgelegten Startwerte werden in Abschnitt 4.2 definiert. Ein allgemeiner Vergleich von Startwerten in Mischverteilungsmodellen ist durch Karlis und Xekalaki (2003) gegeben.

Im folgenden Kapitel werden die vorgestellten Fallzahlanpassungsstrategien gegenübergestellt und detailliert auf ihre Funktionsweise untersucht. Im Fokus stehen die Eigenschaften der Verfahren in Bezug auf die Wahrscheinlichkeit eines Fehler 1. Art, auf die gewünschte Power und auf die Fallzahlverteilung.

KAPITEL 4

ROBUSTHEITSANALYSE VERBLINDER UND UNVERBLINDER FALLZAHLANPASSUNGSSTRATEGIEN

4.1 Vergleich der Prozeduren im Poisson-Modell

4.1.1 Einleitung

Kapitel 4 fokussiert die Prozeduren zur Fallzahlanpassung im Poisson-Modell im adaptiven Design mit interner Pilotstudie (vgl. Abschnitt 3.2). Analytisch und in Simulationsstudien wird überprüft, ob die unverblindeten Strategien eine höhere Effizienz hinsichtlich der Fallzahlschätzung gegenüber verblindeten Verfahren aufweisen. Als Kriterien werden die Wahrscheinlichkeit eines Fehler 1. Arts, die Power und die Fallzahlverteilung herangezogen. Ferner werden die Fallzahlverteilung und die Power, resultierend aus der Anwendung von entblindeten und verblindeten Strategien, für den Fall $\beta_1^* \neq \beta_1$ untersucht. Verblindete wie auch unverblindete störparameterbasierte Fallzahlanpassungsprozeduren sind abhängig von der richtigen Spezifizierung des Effektes. Demzufolge ist die Robustheit der Prozeduren, besonders bei kleinen Abweichungen vom spezifizierten Effekt, für die praktische Anwendung von hohem Interesse. Abschließend werden an zwei randomisierten klinischen Studien in multipler Sklerose von Rudick et al. (2006) und Comi et al. (2012) die Ergebnisse illustriert.

Bevor im Folgenden die einzelnen Merkmale im Detail untersucht werden, wird das für die Simulationsdurchläufe festgelegte Setting vorgestellt. Sofern nicht anders erwähnt, werden $n_{\text{sim}} = 10.000$ Wiederholungen für jedes Szenario durchgeführt. Die Parametereinstellungen sind durch das Krankheitsbild der multiplen Sklerose motiviert und werden in Tabelle 4.1 zusammengefasst.

Tabelle 4.1: Betrachtete Szenarien der Simulationsstudie

	Poissondaten	Dispensierte Daten
Signifikanzniveau: α	0.025 (einseitig)	0.025(einseitig)
Gewünschte Power: $1 - \beta$	0.9	0.9
Anteil π im IPS	0.5	0.5
Allokationsverhältnis k	1	1
Gesamtrate: λ	0.5, 0.51, ..., 1	0.75
Dispersionsparameter: σ^2	1	1.5, 2, 2.5
Angenommene Gesamtrate: λ^*	0.75	0.75
Angenommener Dispersionsparameter: σ^{2*}	1	1
Angenommener Effekt: $\exp(\beta_1^*)$	0.6, 0.75	0.75
Wahrer Effekt: $\exp(\beta_1)$		
unter H_0	1	1
unter H_1	0.45, 0.55, ..., 0.85	0.45, 0.55, ..., 0.85

4.1.2 Fehler 1. Art und Powervergleich

Zunächst wird die Wahrscheinlichkeit für einen Fehler 1. Art für die Fallzahlenanpassungsstrategien im Poisson-Regressionmodell untersucht. Die Kontrolle der Wahrscheinlichkeit für einen Fehler 1. Art ist in klinischen Studien von besonderem Interesse, da der Fehler 1. Art als Konsumentenrisiko gilt (vgl. Kapitel 1). Bestehende Ergebnisse nach Friede und Schmidli (2010a) zeigen, dass das nominale Niveau der verblindeten Fallzahlenanpassungsprozedur (vgl. Abschnitt 3.2.1) eingehalten wird. Für den unverblindeten Ansatz (vgl. Abschnitt 3.2.2) liegen bislang keine Ergebnisse vor. Zur Bewertung der Prozeduren wird auf eine Simulationsstudie zurückgegriffen; die dazugehörigen Szenarien sind in Tabelle 4.1 (Poissondaten) aufgelistet. Analog kann über eine exakte Berechnung der Wahrscheinlichkeit eines Fehlers 1. Art eine Aussage getroffen werden. Eine entsprechende Formel ist von Friede und Schmidli (2010a) und Schneider et al. (2013a) bereitgestellt. Für die folgenden Simulationen wird vorerst die Annahme $\beta_1^* = \beta$ getroffen. Des Weiteren wird die approximative Fallzahlformel aus Gl. 2.14 für die initiale Fallzahlplanung verwendet.

In der Planungsphase der klinischen Studie wird von einer Gesamtrate von $\lambda^* = 0.75$ und dem wahren sowie angenommenen Effekt $\exp(\beta_1^*) = \exp(\beta_1) = 0.6$ oder $\exp(\beta_1^*) = 0.75$ ausgegangen. Entsprechende Parameterannahmen führen zu einer initialen (fixen) Kontrollgruppenfallzahl von $m_0 = 115$ bzw. $m_0 = 346$, damit ein tatsächlich vorliegender Effekt mit einer 90 %-igen Wahrscheinlichkeit aufgedeckt wird. In $n_{\text{sim}} = 100.000$ Wiederholungen pro Simulation werden nun 102 Szenarien betrachtet, in denen die Gesamtrate λ in 0.01 Schritten von $\lambda = 0.5$ bis $\lambda = 1$ ansteigt. Folglich wird die Annahme von $\lambda^* = 0.75$ nur in zwei der 102 Szenarien (für zwei Effektgrößen) zu Beginn richtig spezifiziert.

Die Auswirkungen der Missspezifikationen auf die Fallzahlgröße sind bedeutend. Im ungünstigsten Fall, in dem die Gesamtrate während der klinischen Studie 0.5 beträgt, wird eine Fallzahl von $m_0 = 172$, (519) im Vergleich zu $m_0 = 115$, (346) benötigt, damit die gewünschte Power von 90 %

bei einem Effekt von $\exp(\beta_1^*) = \exp(\beta_1) = 0.6$ (0.75) beibehalten wird. Die Simulationsergebnisse der 102 Szenarien sind in Abbildung 4.1 dargestellt. Bis auf wenige Ausreißer zeigt die verblindete Fallzahlanpassungsprozedur keine Inflation des nominalen Signifikanzniveaus. Im Gegensatz dazu neigt der unverblindete Ansatz zu einem liberaleren Verhalten. Besonders bei größeren Effekten $\exp(\beta_1^*) = 0.6$, welche zu einer geringeren Fallzahl führen, treten liberale Resultate auf. Zum Teil können diese Besonderheiten durch den Waldtest erklärt werden, der bei kleinen Stichprobengrößen bekanntermaßen zu liberalen Ergebnissen neigt. In solchen Szenarien sollte über andere Testprozeduren nachgedacht werden. Konietschke et al. (2014) schlagen einen Permutationsansatz vor, der zu verbesserten Ergebnissen hinsichtlich der Kontrolle des Signifikanzniveaus führt.

Die ermittelten Ergebnisse stimmen mit Resultaten für normalverteilte Daten überein. Die Arbeiten von Coffey und Muller (1999), Wittes et al. (1999), Denne und Jennison (1999) und Kieser und Friede (2000) zeigen ebenfalls eine marginale Überschreitung des Signifikanzniveaus bei Entblindung.

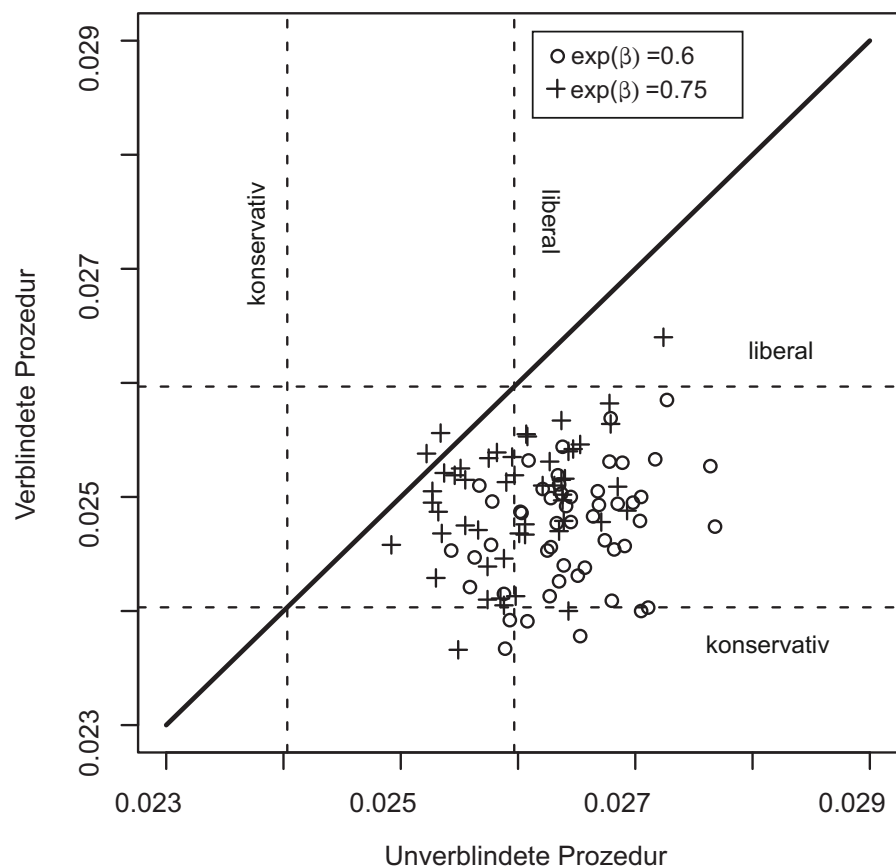


Abbildung 4.1: Fehler 1. Art der 102 simulierten Szenarien von der unverblindeten Prozedur aufgetragen gegen die Resultate der verblindeten Prozedur. Die simulierten Szenarien sind in Tabelle 4.1 zusammengefasst, wobei $\pi = 0.5$ und $n_{\text{sim}} = 100.000$ sind. Die gestrichelten Linien zeigen den erwarteten Streubereich für 95 % beobachtete Häufigkeiten bei 100.000 simulierten Studien und einem tatsächlichen Fehler 1. Art von $\alpha = 0.025$.

Durch die Fallzahlanpassungsstrategien wird eine Sicherung der vorsepezifizierten Power unter bestimmten Umständen gewährleistet. Im Folgenden werden mit den gleichen Annahmen, die zur Niveausimulation genutzt wurden, entsprechende Powersimulationen durchgeführt. Beide Prozeduren liefern zufriedenstellende Ergebnisse. Für einen Effekt von $\exp(\beta_1) = 0.75$ bzw. $\exp(\beta_1) = 0.6$ liegt der Mittelwert [Konfidenzintervall] der 51 Szenarien für die verblindete Prozedur bei 0.904 [0.902, 0.906] bzw. 0.913 [0.911, 0.915]. Die Resultate der unverblindeten Prozedur sind durch 0.921 [0.919, 0.922] für $\exp(\beta_1) = 0.75$ bzw. durch 0.938 [0.935, 0.940] für einen Effekt von $\exp(\beta_1) = 0.6$ gegeben. Zusammenfassend wird deutlich, dass die unverblindete Prozedur zu einer höheren Power führt als gewünscht ist. Dieses Phänomen ist bedingt durch das liberale Verhalten unter der Nullhypothese und durch die größere Variabilität der Schätzer; beide Aspekte werden im folgenden Abschnitt diskutiert.

4.1.3 Vergleich der Fallzahlverteilungen

Zugunsten einer besseren Lesbarkeit wird im Folgenden die Annahme $t_{ij} = t = 1$ an die Nachbarbeobachtungszeiten der Patienten gestellt. Die in diesem Abschnitt gezeigten Resultate können unter Voraussetzung 2.1 verallgemeinert werden. Zunächst wird auf den verblindeten und unverblindeten Schätzer der Kontrollrate zurückgegriffen. Mit der Annahme $t_{ij} = t = 1$ und einem Randomisierungsverhältnis von $1 : k$ ($m_0 = k m_1$) vereinfachen sich die Schätzer aus Gl. 3.5 und Gl. 3.6 zu

$$\hat{\lambda}_{0(b)} = \frac{1}{m_0^*(k+1)} \left(\sum_{i=0}^1 \sum_{j=1}^{m_i^*} n_{ij} \right) \left(\frac{k+1}{k \exp(\beta_1^*) + 1} \right) \quad (4.1)$$

und

$$\hat{\lambda}_{0(u)} = \frac{1}{m_0^*} \sum_{j=1}^{m_0^*} n_{0j}. \quad (4.2)$$

Die Schätzer $\hat{\lambda}_{0(b)}$ und $\hat{\lambda}_{0(u)}$ folgen mit der Voraussetzung 2.1 asymptotisch einer Normalverteilung (vgl. Satz 2.2), wobei der verblindete Schätzer von einer Effektspezifikation abhängig ist. Die Verteilungen sind durch die Ausdrücke

$$\begin{aligned} \sqrt{m_0^*} \left[\hat{\lambda}_{0(b)} - \lambda_0 \left(\frac{k \exp(\beta_1) + 1}{k \exp(\beta_1^*) + 1} \right) \right] &\stackrel{\mathcal{D}}{\sim} \mathcal{N} \left(0, \lambda_0 \left(\frac{1}{k \exp(\beta_1^*) + 1} \right)^2 (k \exp(\beta_1) + 1) \right) \\ \sqrt{m_0^*} [\hat{\lambda}_{0(u)} - \lambda_0] &\stackrel{\mathcal{D}}{\sim} \mathcal{N} (0, \lambda_0) \end{aligned}$$

gegeben.

Die Erwartungstreue des verblindeten Schätzers ist für eine inkorrekte Effektannahme $\beta_1^* \neq \beta_1$ nicht mehr gegeben. Zur Bewertung der Schätzer in Gl. 4.1 und Gl. 4.2 wird die mittlere quadratische Abweichung (MSE) herangezogen, die sich aus der Varianz und der quadratischen Verzerrung der Schätzer zusammensetzt. Der MSE des unverblindeten Schätzers ist durch die Eigenschaft der Er-

wartungstreue aus Satz 3.3 mit der Varianz gleichzusetzen:

$$\text{MSE}(\hat{\lambda}_{0(u)}) = \frac{\lambda_0}{m_0^*} \quad (4.3)$$

Für den verblindeten Schätzer wird die Verzerrung zu dem Ausdruck $\lambda_0 \frac{k(\exp(\beta_1) - \exp(\beta_1^*))}{(k \exp(\beta_1^*) + 1)}$ berechnet, so dass der MSE durch

$$\text{MSE}(\hat{\lambda}_{0(b)}) = \frac{\lambda_0}{m_0^*} \frac{(k \exp(\beta_1) + 1)}{(k \exp(\beta_1^*) + 1)^2} + \lambda_0^2 \frac{k^2 (\exp(\beta_1) - \exp(\beta_1^*))^2}{(k \exp(\beta_1^*) + 1)^2} \quad (4.4)$$

gegeben ist. Für die Annahme $\beta_1^* = \beta_1$ ist der mittlere quadratische Fehler der verblindeten Prozedur im Vergleich zur unverblindeten Prozedur um den Faktor $1/(1 + k\beta_1)$ kleiner. In einer klinischen Studie mit einem wahren Effekt von $\exp(\beta_1) = 0.6$ und balancierter Randomisierung ist der Faktor durch 0.625 gegeben. Wird die Behandlungsgruppe doppelt gewichtet ($k = 2$), so ist die mittlere quadratische Abweichung durch die Anwendung eines verblindeten Verfahrens schon mehr als halbiert ($1/(0.6 \cdot 2 + 1) = 0.46$). In einem weiteren Schritt wird analysiert, bis zu welcher Fehleinschätzung von β_1 diese Überlegenheit erhalten bleibt. Dazu werden - durch Gleichsetzung von Gl. 4.3 und Gl. 4.4 - zwei Grenzen (d_L und d_U) abhängig vom wahren Effekt $\exp(\beta_1)$ berechnet, die zu einem identischen MSE beider Schätzer führen. Die Grenzen sind durch folgende Ausdrücke definiert:

$$d_L = \left(\exp(\beta_1^*) - \frac{1}{2m_0^* \lambda_0} \right) - \frac{\sqrt{4 \exp(\beta_1^*) \lambda_0 m_0^* (1 + \exp(\beta_1^*)) + 1}}{2 \lambda_0 m_0^*} \quad (4.5)$$

$$d_U = \left(\exp(\beta_1^*) - \frac{1}{2m_0^* \lambda_0} \right) + \frac{\sqrt{4 \exp(\beta_1^*) \lambda_0 m_0^* (1 + \exp(\beta_1^*)) + 1}}{2 \lambda_0 m_0^*} \quad (4.6)$$

Besteht eine große Unsicherheit in der Effekteinschätzung, so kann eine unverblindete Fallzahlanpassungsstrategie hinsichtlich der mittleren quadratischen Abweichung von Vorteil sein. Diese Beziehung wird im Abschnitt 4.1.4 an zwei Beispielen illustriert. Da die Schätzer zusätzlich eine direkte Auswirkung auf die geschätzte Fallzahl haben, können diese Ergebnisse in der Planungsphase einer klinischen Studie benutzt werden, um mögliche Auswirkungen einer Fehleinschätzung zu quantifizieren und qualifizieren. Die Konsequenz für die Fallzahlverteilung wird im Folgenden analysiert.

In den Gleichungen 4.1 und 4.2 werden die Schätzer in die Fallzahlformel in Gl. 2.14 eingesetzt, so dass die resultierenden Fallzahlen $\hat{m}_{0(u)}$ oder $\hat{m}_{0(b)}$ erneut als Schätzer interpretiert werden können. Die approximativen Verteilungen der Fallzahlsschätzer werden unter Anwendung von Satz A.1 zu

$$\sqrt{m_0^*} [\hat{m}_{0(u)} - m_0] \stackrel{\mathcal{D}}{\sim} \mathcal{N} \left(0, g(\alpha, \beta, \exp(\beta_1^*), k)^2 \frac{1}{\lambda_0^3} \right) \quad (4.7)$$

für die unverblindete und

$$\sqrt{m_0^*} \left[\hat{m}_{0(b)} - m_0 \left(\frac{k \exp(\beta_1^*) + 1}{k \exp(\beta_1) + 1} \right) \right] \stackrel{\mathcal{D}}{\sim} \mathcal{N} \left(0, g(\alpha, \beta, \exp(\beta_1^*), k)^2 \frac{1}{\lambda_0^3} \frac{(k \exp(\beta_1^*) + 1)^2}{(k \exp(\beta_1) + 1)^3} \right) \quad (4.8)$$

für die verblindete Fallzahlanpassungsstrategie berechnet, wobei die Funktion $g(\alpha, \beta, \exp(\beta_1^*), k) = (z_{1-\beta} + z_{1-\alpha})^2 / (\beta_1^*)^2 (1 / (k \exp(\beta_1^*)) + 1)$ ist. Zu beachten ist, dass die unverblindete sowie die verblindete Fallzahlverteilung von β_1^* abhängen. Diese Eigenschaft resultiert aus der Idee der störparameterbasierten Fallzahlanpassung, die keine Einschätzung der Effektgröße vorsieht. Zur Evaluierung der Prozeduren werden folgende Fälle unterschieden; die Bewertung ist dabei abhängig von der analytisch bestimmten Verzerrung und Varianz der Fallzahlverteilung:

1. $\exp(\beta_1) = \exp(\beta_1^*)$: Liegt eine korrekte Effekteinschätzung vor, so liefern beide Ansätze asymptotisch die benötigte Fallzahl m_0 (vgl. Gl. 4.7 und Gl. 4.8). Im Varianzvergleich stellt sich heraus, dass die Varianz der unverblindeten Prozedur im Vergleich zur verblindeten um den Faktor $(k \exp(\beta_1) + 1)$ größer ist. Bezogen auf das Beispiel (siehe MSE), ist die Varianz der unverblindeten Fallzahlanpassung bei einem wahren Effekt von $\exp(\beta_1) = 0.6$ z. B. um den Faktor 1.6 größer.
2. $\exp(\beta_1) < \exp(\beta_1^*)$: Ist der beobachtete Effekt größer als der angenommene Effekt, so wird keine Veränderung der Fallzahlverteilung bei Entblindung festgestellt. Das unverblindete Verfahren führt unabhängig vom wahren Effekt asymptotisch zur Fallzahl m_0 . Die Fallzahlschätzung ist hinsichtlich der wahren Effektgröße keineswegs optimal, jedoch im Kontext der störparameterbasierten Fallzahlanpassung. Die benötigte Fallzahl $m_{0(req)}$ wird für $\exp(\beta_1) < \exp(\beta_1^*)$ kleiner sein als der Erwartungswert m_0 .

Für $\exp(\beta_1) < \exp(\beta_1^*)$ ist eine systematische Verzerrung für die verblindete Prozedur zu erkennen, die durch die Verzerrung in der Ratenschätzung erklärbar ist. Die vorliegende Fehleinschätzung ist durch den Faktor $(k \exp(\beta_1^*) + 1) / (k \exp(\beta_1) + 1)$ quantifizierbar. Demzufolge führt die Unterschätzung der Rate in Erwartung zu einer Erhöhung der Fallzahl, was im Widerspruch zur benötigten Stichprobengröße steht. Zusammenfassend kann die logische Anordnung $E(\hat{m}_{0(b)}) \geq E(\hat{m}_{0(u)}) \geq m_{0(req)}$ aufgestellt werden. Inwieweit die Unterschiede zwischen den erwarteten Fallzahlen $E(\hat{m}_{0(b)})$ und $E(\hat{m}_{0(u)})$ und der benötigten Fallzahl $m_{0(req)}$ relevant sind, wird im Anschluss an einer Simulationsstudie verdeutlicht.

3. $\exp(\beta_1) > \exp(\beta_1^*)$: Ist der wahre Effekt kleiner als zunächst angenommen, so bleibt die erwartete Fallzahl bei Entblindung konstant. Dagegen wird die Fallzahl bei Verblindung in Erwartung herabgesetzt. Auch in diesem Fall lässt sich eine logische Anordnung aufzeigen: $E(\hat{m}_{0(b)}) \leq E(\hat{m}_{0(u)}) \leq m_{0(req)}$.

Um die Variabilität der Schätzung einzubeziehen, werden die Varianzen der Fallzahlverteilungen in Gleichung 4.7 und 4.8 gleichgesetzt und resultieren im Ausdruck

$$\exp(\beta_1) = \frac{(k \exp(\beta_1^*) + 1)^{2/3} - 1}{k}. \quad (4.9)$$

Für den Fall $k = 1$ lässt sich der Ausdruck in Gl. 4.9 durch $\exp(\beta_1) \approx 0.6 \exp(\beta_1^*)$ approximieren. Das bedeutet, dass für $\exp(\beta_1) < 0.6 \exp(\beta_1^*)$ die Varianz der verblindeten Fallzahlverteilung im Vergleich zur unverblindeten kleiner ist. Das verblindete Verfahren liefert also bis zu einer Effekt-Fehleinschätzung von 40 % eine kleinere Variabilität der geschätzten Fallzahl.

Die dargestellten Fallzahlverteilungen und zusätzlichen Resultate über die Verzerrung und Variabilität werden im Folgenden mit Hilfe einer Simulationsstudie auf ihre Exaktheit überprüft. Dazu wird ein spezielles Szenario aus Tabelle 4.1 ($\lambda^* = 0.75$ und $\exp(\beta_1^*) = 0.75$) herangezogen. Sind die Annahmen in der Planung korrekt gewählt, so ergibt sich eine initiale Fallzahl von $m_0 = 346$ für die Kontrollgruppe. Im Folgenden werden die verschiedenen Missspezifikationen der Effektgröße (1.-3.) betrachtet und untersucht. Zusätzlich wird von einer Fehleinschätzung der Gesamtrate ausgegangen. Die Resultate über die Fallzahlverteilung und die Powereigenschaften bei Ent- oder Verblindung sind in Tabelle 4.2 zusammengefasst. In den Spalten „num“ sind die numerisch bestimmten Werte der Fallzahl aufgeführt.

Fallzahlverteilung und Variabilität

Die Herleitung der Approximationen der Fallzahlschätzung liefert angemessene Resultate, die mit den simulierten Werten übereinstimmen (Tabelle 4.2). Zunächst wird der Fall $\exp(\beta_1) = \exp(\beta_1^*)$ betrachtet. Beide Prozeduren liefern asymptotisch die erwartete Fallzahl m_0 . Da ein balanciertes Design vorliegt, unterscheiden sich die Standardabweichungen der Fallzahlverteilungen um den Faktor $\sqrt{\exp(\beta_1) + 1} = 1.32$. Der unverblindete Ansatz weist demzufolge eine erheblich größere Streuung in der Fallzahlverteilung auf. Bei der Betrachtung der Szenarien, in denen eine Effektunsicherheit ($\exp(\beta_1) < (>) \exp(\beta_1^*)$) vorliegt, werden die zuvor erläuterten Eigenschaften für die Prozeduren beobachtet. Bei Entblindung wird keine Veränderung der asymptotischen Fallzahlverteilung registriert. Demgegenüber weist der verblindete Ansatz einerseits unnötig hohe Fallzahlen bei wachsendem Effekt und andererseits zu kleine Stichproben bei abnehmender Effektgröße auf.

Wird der Zusammenhang der Varianzen in Gleichung 4.9 berücksichtigt, so wird mit der Annahme $\exp(\beta_1^*) = 0.75$ eine Effektgröße von $\exp(\beta_1) = 0.45$ berechnet. Dieser Wert gibt die Größe an, bei dem das Streuungsverhältnis der Verteilungen äquivalent ist. Eine Unterschreitung der Grenze, führt zu dem Resultat $SD(\hat{m}_{0(b)}) \geq SD(\hat{m}_{0(u)})$ (Tabelle 4.2). Demzufolge ist die Streuung der verblindeten Fallzahlverteilung bis zu einem vorliegenden Effekt von $\exp(\beta_1) = 0.45$ im Vergleich zur entblindeten Verteilung geringer.

Vergleich für die benötigte Fallzahl $m_{0(req)}$

Im Folgenden werden die resultierenden Fallzahlen der Prozeduren mit der benötigten Größe $m_{0(req)}$ für einen wahren Effekt $\exp(\beta_1)$ verglichen. Beide Ansätze zur Fallzahlanpassung sind hinsichtlich der Power nicht robust, wenn die Effektgröße falsch spezifiziert ist (Tabelle 4.2). Für einen größeren Effekt als angenommen wird eine kleinere Stichprobe $m_{0(req)}$ benötigt, damit der Unterschied mit einer Power von $1 - \beta$ gezeigt werden kann. Demzufolge stehen die Resultate der Anpassungsproze-

duren im Widerspruch zu den abnehmenden Fallzahlen bei wachsendem Effekt. Liegt ein geringerer Effekt vor, so erreichen beide Verfahren die gewünschte Power nicht, da sie in der Fallzahl konstant bleiben beziehungsweise diese verringern. Was sind die Konsequenzen für die praktische Anwendung?

In einer Studienplanung kann der *minimal* klinisch relevante Effekt zur Berechnung der initialen Fallzahl verwendet werden. Die Wahl eines minimalen Effekts beinhaltet die folgenden Auswirkungen. Liegt ein geringerer Effekt vor als angenommen, so erreichen beide Verfahren die gewünschte Power nicht. Diese Eigenschaft ist von geringerer Bedeutung in der konservativen Studienplanung, da ein geringerer Effekt unter der klinischen Relevanzgrenze liegt. Ist der beobachtete Effekt in der laufenden klinischen Studie größer als angenommen, wird die Fallzahl lediglich erhöht. Demzufolge ist eine Powersteigerung auf Kosten einer Ressourcenverschwendung zu erwarten, wodurch der Erfolg der Studie jedoch nicht beeinträchtigt ist.

Zusammenfassend wird festgestellt, dass beide vorgestellten Fallzahlanpassungsstrategien die gewünschten Eigenschaften hinsichtlich Power und Kontrolle der Fehlerrate unter $\beta_1 = \beta_1^*$ besitzen. Der entblindete Ansatz zeigt jedoch leicht liberale Ergebnisse und zudem zu hohe Power der Studien. Existiert eine Unsicherheit in der Effektgröße, so führen beide Verfahren zu Fehlschätzungen der benötigten Fallzahl, die jedoch in konservativen Studienplanungen zu keinem Misserfolg führen. Wird zusätzlich die praktische Relevanz der Verzerrung bei Verblindung betrachtet, so sind die auftretenden Differenzen der erwarteten Fallzahlen $E(\hat{m}_{0(b)}) - E(\hat{m}_{0(u)})$ vernachlässigbar, wenn die tatsächlich benötigte Stichprobe $m_{0(req)}$ hinzugezogen wird (Tabelle 4.2). Im Hinblick auf die oftmals geringere Streuung der Fallzahlschätzung ist eine verblindete Fallzahlanpassung in vielen Situationen nicht nur aus regulatorischer Sicht zu empfehlen. Im Folgenden werden die Ergebnisse an zwei Studien illustriert.

4.1.4 Beispiel

Die Resultate werden an zwei randomisierten, kontrollierten und doppel-verblindeten klinischen Studien im Krankheitsbild der multiplen Sklerose verdeutlicht.

In der Studie nach Rudick et al. (2006) erhalten 589 Patienten eine Kombinationstherapie und 582 eine alleinige Interferon beta-1a Behandlung. Die Patienten wurden zufällig in diese Behandlungsarme randomisiert. Primäre Endpunkte waren die Schubrate nach einem Jahr und der Behinderungsgrad nach zwei Jahren. Die jährliche Schubrate ergab sich für die Kombinationstherapie zu 0.38 und für den weiteren Behandlungsbereich zu 0.81 (vgl. Tabelle 4.3). In Tabelle 1 in Friede und Schmidli (2010a) wurde gezeigt, dass eine Poissonverteilung eine geeignete Beschreibung für die beobachteten Häufigkeiten ist. Demzufolge ist die angewandte Poisson-Regression ein adäquates Verfahren zur Schätzung des finalen Behandlungseffektes, der zudem ein signifikantes Ergebnis zugunsten der Kombinationstherapie zeigt.

Tabelle 4.2: Merkmale der geschätzten Fallzahlverteilung illustriert an einem Szenario aus Tabelle 4.1 bei Entblindung (entb.) bzw. Verblindung (verb.). Zusätzlich ist die benötigte Fallzahl für den korrespondierenden Effekt $\exp(\beta_1)$ gegeben. Der angenommene Effekt beträgt $\exp(\beta_1^*) = 0.75$.

			Mittlere Fallzahl				SD der Fallzahl				Power	
			verb.		entb.		verb.		entb.		verb.	entb.
λ	e^{β_1}	$m_{0(req)}$	sim	num	sim	num	sim	num	sim	num		
0.5	0.45	92.9	629.8	625.7	523.7		53.3	52.3	53.6		100	100
	0.55	145.0	589.1	585.3	523.8		48.2	47.3	53.8		100	100
	0.65	251.5	553.3	549.8	523.7	518.4	43.8	43.1	53.7	52.1	99.8	99.8
	0.75	518.4	521.4	518.4	523.6		40.0	39.4	53.8		90.4	91.93
	0.85	1515.2	493.0	490.4	523.4		36.8	36.3	53.7		45.8	48.5
0.75	0.45	61.9	419.1	417.1	347.9		28.7	28.4	28.9		100	100
	0.55	96.7	392.0	390.2	348.1		26.2	25.7	29.0		100	100
	0.65	167.7	368.1	366.6	347.9	345.6	23.7	23.4	29.0	28.4	99.8	99.9
	0.75	345.6	346.9	345.6	348.1		21.7	21.5	29.0		90.4	91.7
	0.85	1010.2	328.2	326.9	347.8		19.9	19.7	28.8		45.7	48.6
1	0.45	46.5	313.9	312.8	260.5		18.7	18.5	18.6		100	100
	0.55	72.5	293.6	292.7	260.3		16.9	16.7	18.7		100	100
	0.65	125.8	275.8	274.9	260.5	259.2	15.4	15.2	18.7	18.4	99.8	99.9
	0.75	259.2	259.9	259.2	260.3		14.0	13.9	18.6		90.7	92.0
	0.85	757.6	245.9	245.2	260.1		12.9	12.8	18.7		46.3	48.3

Tabelle 4.3: Eigenschaften der motivierenden klinischen Studien, wobei RR das Ratenverhältnis, T_R die Rekrutierungsperiode und ARR die jährliche Schubrate sind. Ein Index „obs (as)“ deutet auf einen beobachteten (angenommenen) Wert hin.

Klinische Studie	Behandlungsarm	n	T_R [Monate]	ARR_{obs}	ARR_{as}	RR_{obs}	RR_{as}	$\sigma_{i(obs)}^2$	$\sigma_{i(as)}^2$
Rudick et al. (2006)	Add-on Placebo Kombinationstherapie	582 589	12	0.81 0.38	0.9 -	0.47	0.66	-	-
Comi et al. (2012)	Placebo Oral laquinimod	556 550	12	0.39 0.30	0.65 -	0.77	0.75	1.6 1.8	1.6

Es ist erkennbar, dass die beobachtete Kontrollrate $\hat{\lambda}_0 = 0.81$ kleiner ist als die angenommene Kontrollrate $\lambda_0 = 0.9$; somit hätte die Studie von einer Fallzahlanpassungsstrategie profitiert. Im Folgenden wird angenommen, dass die Zwischenauswertung vergleichsweise früh in der Studie stattfindet - z. B. wenn 120 Patienten eingeschlossen sind (20 % von der initialen Fallzahl). Dieser Ansatz ist bedingt durch die relativ schnelle Rekrutierung von 12 Monaten plausibel. Die Parameter d_L und d_U der verblindeten und entblindeten Prozedur werden mit Hilfe des angenommenen Effekts, der Fallzahl zur Zwischenauswertung und der wahren Kontrollrate durch Gl. 4.5 bzw. Gl. 4.6 ermittelt. Diese definieren den wahren Effekt β_1 , für den der mittlere quadratische Fehler beider Schätzer der Kontrollrate äquivalent ist. Die Berechnung ergibt eine untere Grenze $d_L = 0.55$. Da der beobachtete Effekt mit $\exp(\beta_1) = 0.47$ unterhalb des Wertes liegt, ist der mittlere quadratische Fehler des unverblindeten Kontrollratenschätzers im Vergleich zum verblindeten kleiner. Abbildung 4.2 zeigt, dass sich dieses Verhalten nur für sehr kleine Fallzahlen wendet ($m_0^* = 40$) und erst dann die verblindete

Prozedur Vorteile aufzeigt. In allen anderen Fällen (besonders für spätere Zwischenauswertungen) wäre ein unverblindeter Ansatz zur Fallzahlanpassung vorzuziehen. Diese Eigenschaften lassen sich durch die erhebliche Unsicherheit hinsichtlich der Effektgröße erklären. Der angenommenen Effekt weicht um 41 % vom tatsächlich beobachteten Effekt (0.47) ab. In Situationen, in denen die Effektstärke sehr variabel ist, sollten andere Designs wie gruppensequentielle Pläne, die eine frühere Studienbeendigung vorsehen, vorteilhaft sein (Whitehead, 1997; Jennison und Turnbull, 2000).

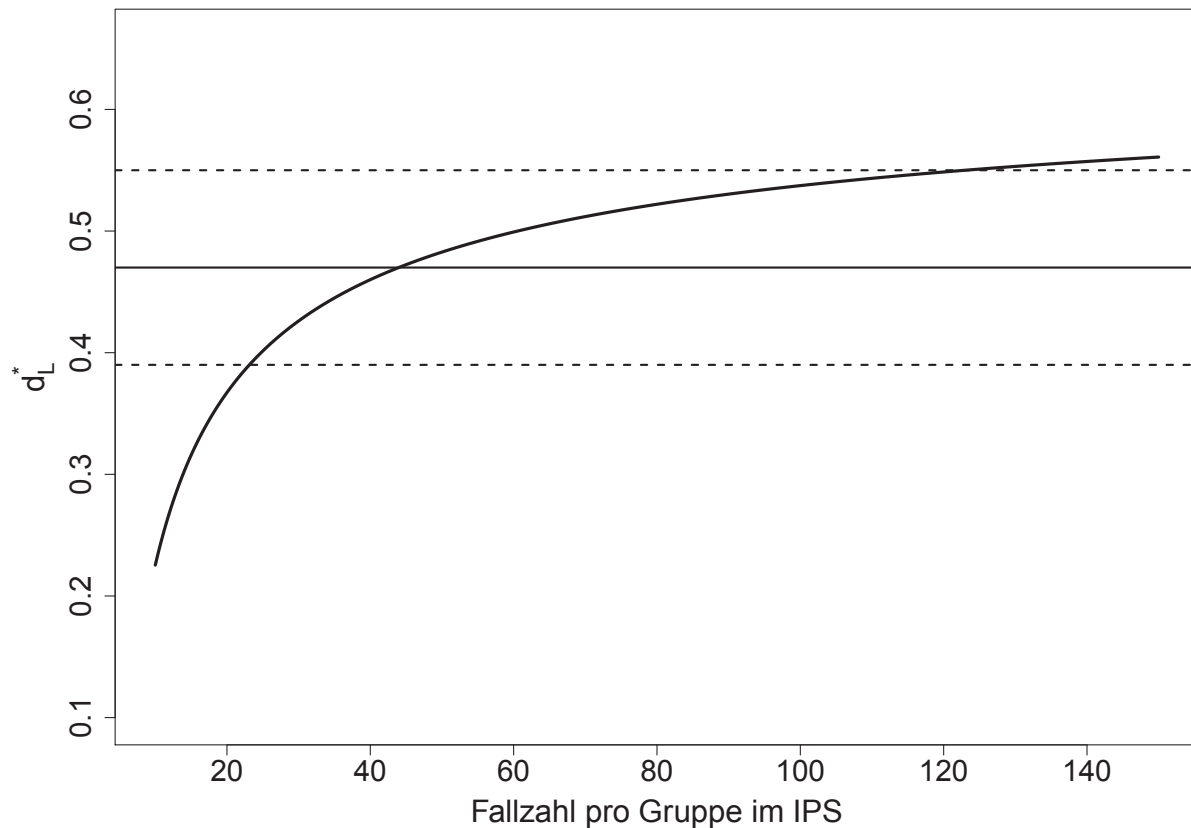


Abbildung 4.2: Darstellung der Grenzwerte d_L , an denen der mittlere quadratische Fehler der ver- und entblindeten Prozedur in Abhängigkeit der Fallzahl m_0 gleich ist. Die gestrichelten Linien stellen das 95 % Konfidenzintervall der wahren Effektgröße $\exp(\beta_1) = 0.47$ dar. Das Intervall wurde mit Hilfe der Daten nach Rudick et al. (2006) berechnet.

Die randomisierte, doppel-verblindete und Placebo-kontrollierte 2-Jahres-Studie von Comi et al. (2012) fand in 24 Ländern mit 1109 Patienten statt. Die Phase III Studie besaß eine Rekrutierungsperiode von 12 Monaten, in der die Patienten zufällig in eine Behandlungsgruppe (orale Gabe von Laquinimod 0.6 mg pro Tag) und eine Placebo-Gruppe eingeteilt wurden. Der primäre Endpunkt war dabei die jährliche Schubrate der Patienten. Um den Endpunkt am Ende der Studie zu analysieren, wurde ein Quasi-Likelihood Ansatz, der eine Überdispersion zulässt, gewählt. Die Annahme über die Ungleichheit von Erwartungswert und Varianz scheint gerechtfertigt (Tabelle 4.3). Ein formaler Test auf Überdispersion, der z. B. von Xing und Lee (2005) vorgeschlagen wird, wurde nicht durchgeführt. Nach Ablauf der 24 Monate konnte eine adjustierte Behandlungs- bzw. Kontrollrate

von 0.3 und 0.39 geschätzt werden (Tabelle 4.3).

Die Berechnung des Überdispersionsschätzers der beiden Gruppen erfolgte mit Hilfe der korrespondierenden Standardfehler (entnommen aus Comi et al. (2012)). Die Schätzer sind durch 1.6 für die Behandlungsgruppe bzw. 1.8 für die Kontrollgruppe gegeben.

In der Planungsphase der klinischen Studie wurde von einer Kontrollrate von $\lambda_0^* = 0.65$ und einem klinisch relevanten Effekt von $\exp(\beta_1^*) = 0.75$ ausgegangen. Obwohl keine Aussage zur Überdispersion getätigt wurde, liegt eine Annahme für den Dispersionsparameter von $\sigma^{2*} = 1.8$ nahe, damit die initiale Fallzahlschätzung von Comi et al. (2012) reproduzierbar ist. Die tatsächlich beobachtete Kontrollrate wird mit $\hat{\lambda}_0 = 0.39$ angegeben, was bei einem Effekt von $\exp(\beta_1) = 0.77$ zu einer drastischen Reduzierung der Power führt (Comi et al., 2012). Um eine Power von 90 % zu sichern, muss die initial berechnete Fallzahl um 66 % erhöht werden. Demzufolge besteht kein Zweifel, dass die Studie von einem Design mit interner Pilotstudie profitiert hätte, obwohl sie letztendlich ein signifikantes Ergebnis zeigte.

Der mittlere quadratische Fehler des verblindeten Schätzers der Kontrollrate ist um den Faktor 0.625 kleiner als der entblindete Schätzer, falls $\exp(\beta_1^*) = \exp(\beta_1)$ und $k = 1$ gilt (vgl. Abschnitt 4.1.3). Da der angenommene Effekt nur um 0.02 vom beobachteten Effekt abweicht, ist diese Annahme gerechtfertigt. Infolgedessen ist die verblindete Prozedur hinsichtlich des mittleren quadratischen Fehlers vorteilhaft. Im Folgenden wird nun die Fallzahlverteilung der Prozeduren näher betrachtet. Dafür wurde die geschätzte Fallzahl in 20.000 Wiederholungen mit den wahren Parameterwerten (Tabelle 4.3) simuliert. Abbildung 4.3 zeigt die Ergebnisse der ver- und entblindeten Prozedur für verschiedene Fallzahlgrößen in der Zwischenauswertung. Für alle sichtbaren Fallzahlgrößen ist die unverblindeten Schätzung breiter verteilt. Besonders zu frühen Zeitpunkten (z. B. $m_0 = 100$) ist die Differenz zwischen den resultierenden Verteilungen beachtlich. Insofern ist es offensichtlich, dass die Anwendung einer verblindeten Fallzahlanpassungsprozedur in der betrachteten klinischen Studie vorteilhaft gewesen wäre.

Abschließend ist festzustellen, dass verblindete und entblindete Verfahren zur Fallzahlanpassung in vielen Situationen zu geeigneten Resultaten in klinischen Studien führen. Jedoch geht die Überlegenheit einer unverblindeten Fallzahlanpassungsprozedur fast immer mit einer drastischen Unsicherheit in der Effektgröße einher.

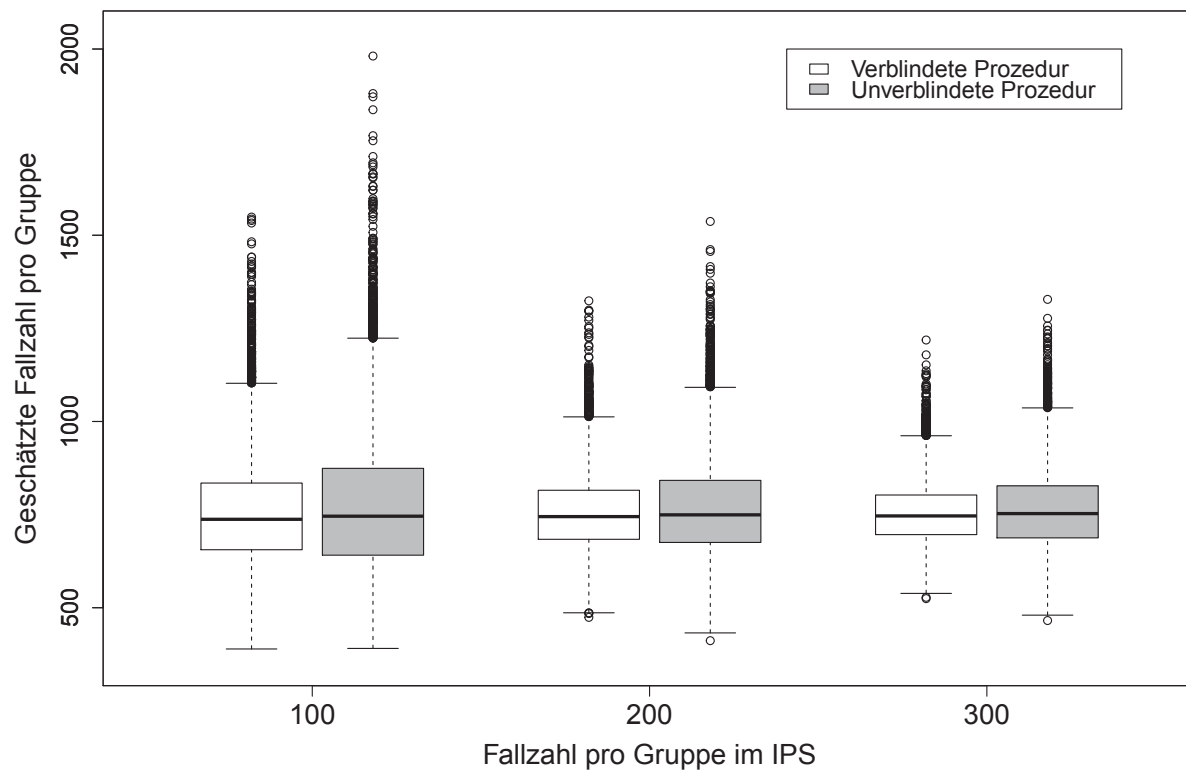


Abbildung 4.3: Fallzahlverteilung der ver- und entblindeten Anpassungsprozedur für verschiedene Fallzahlgrößen in der Zwischenauswertung (20.000 Wiederholungen pro Szenario). Die Boxplots zeigen den Median (horizontale Linie), wobei die Whiskers das 1.5-fache des Interquartilabstandes widerspiegeln.

4.2 Robustheitsanalyse im Negativ-Binomial-Modell

4.2.1 Einleitung

Im Folgenden werden die in Abschnitt 3.3 vorgestellten Verfahren zur Fallzahlanpassung im Negativ-Binomial-Modell betrachtet. Einleitend stellt sich die Frage, warum die Prozeduren nach Friede und Schmidli (2010b) sowie Cook et al. (2009) nach erfolgreicher Publikation auf ihre Funktionsweise erneut untersucht werden. Der Ansatz eines auf einem EM-Algorithmus-basierenden Verfahrens zur Fallzahlanpassung ist kein Novum und wurde erstmalig für die Varianzschätzung bei normalverteilten Daten vorgestellt (Gould und Shih, 1992). Detaillierte Untersuchungen der Prozedur zeigen, dass die Ergebnisse der Maximum-Likelihood Schätzung zu verzerrten und sehr variablen Fallzahlen führen (Friede und Kieser, 2002; Waksman, 2007). Zusätzlich konnte die Abhängigkeit der Resultate von der Initialisierung der Startparameter und vom Konvergenzkriterium gezeigt werden. Erst bei sehr großen Fallzahlen, die keine klinische Relevanz aufweisen, führt der Ansatz zu angemessenen Ergebnissen. Ungeachtet dieser Resultate wurde die auf einem EM-Algorithmus-basierende Methodik zum verblindeten Schätzen der Kontrollrate bei Zähldaten nach Cook et al. (2009) vorge-

schlagen. Aus diesem Grund liegt der Fokus auf dem Vergleich der verblindeten Ansätze. Die Notation wird aus den vorherigen Abschnitte übernommen. Zudem wird die von Friede und Schmidli (2010b) vorgeschlagene Prozedur als „ML-Lump“, die EM-Algorithmus-basierende Prozedur von Cook et al. (2009) als „ML-EM“ und der Quasi-Likelihood Ansatz als „QL-Lump“ bezeichnet. Das Kapitel ist wie folgt strukturiert. Zu Beginn werden die Simulationsszenarien detailliert erläutert. Daran anschließend liegt das Augenmerk auf dem Vergleich zwischen den klassischen Charakteristika der ML-Lump und den jeweiligen Eigenschaften der ML-EM Prozedur. Im Anschluss daran werden die ML- und QL-Lump Prozeduren auf ihre Robustheit hinsichtlich der Heterogenitätsannahme überprüft.

4.2.2 Szenarien der Simulationsstudie

Die Szenarien, die der Simulationsstudie zugrunde liegen, sind in Tabelle 4.4 zusammengefasst. Für den Vergleich zwischen der ML-Lump und der ML-EM Prozedur ist die Spalte „NB-Modell“ von Interesse. Es werden, falls nicht anders erwähnt, $n_{\text{sim}} = 10.000$ Wiederholungen für jedes Szenario simuliert. Hierbei liegt der Fokus auf balancierten Designs ($k = 1$).

Tabelle 4.4: Betrachtete Szenarien in der Simulationsstudie im NB-Modell

	NB-Daten	Robustheit
Signifikanzniveau: α	0.05 (zweiseitig)	0.05 (zweiseitig)
Gewünschte Power: $1 - \beta$	0.8, 0.9	0.8
Anteil π im IPS	0.5	0.5
k	1	1
Konvergenzschranke δ	$10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-8}$	(-)
Kontrollrate: λ_0	2, 4	2, 4
Shapeparameter: ϕ	0.6, 0.9	0.1, 0.2, ..., 3
Angenommene Rate: $\lambda_0^* = x \cdot \lambda_0$	$x \in \{0.5, 0.75, 1, 1.25, 1.5\}$	2, 4
Angenommener Shapeparameter: ϕ	0.6	0.6
Angenommener Effekt: $\exp(\beta_1^*)$	0.75	0.75
Wahrer Effekt: $\exp(\beta_1)$		
unter H_0	1	1
unter H_1	0.75	0.75

In Abschnitt 3.5.2 wurde die Abhängigkeit der resultierenden Parameterschätzer von den gewählten Startwerten und Konvergenzschranken diskutiert. Für das vordefinierte Konvergenzkriterium

$$\max_{1 \leq i \leq 3} \{|\hat{\theta}_{i(k+1)} - \hat{\theta}_{i(k)}|\} < \delta$$

sind in der vorliegenden Arbeit bislang schwache wie auch starke Regeln betrachtet worden ($\delta = 10^{-2}, 10^{-3}, \dots, 10^{-8}$). Um eine Reproduzierung der Ergebnisse von Cook et al. (2009) zu sichern, wird im Folgenden die gleiche Vorgehensweise verwendet. Die Startwerte beschreiben die wahren Parameterwerte θ und eine Unsicherheit, die durch eine normalverteilte Zufallsvariable ausgedrückt

wird und ergeben sich zu:

$$\theta_0 = \theta + \epsilon, \text{ wobei } \epsilon_1, \dots, \epsilon_3 \sim N(0, \sigma) \text{ mit } \sigma = 0.1. \quad (4.10)$$

4.2.3 Fehler 1. Art und Powervergleich

In einem ersten Schritt wird die Konvergenzregel für aufeinanderfolgende Parameterschätzer auf $\delta = 10^{-8}$ gesetzt, um eine Konvergenz des EM-Algorithmus zu sichern. Die maximale Wahrscheinlichkeit für einen Fehler 1. Art beträgt für alle simulierten Szenarien (Tabelle 4.4) 0.0055. Bei einem Monte-Carlo Fehler von 0.002 für $n_{\text{sim}} = 10.000$ Wiederholungen kann von keiner systematischen Inflation des nominalen Signifikanzniveaus der beiden Prozeduren (ML-Lump und ML-EM) gesprochen werden. Übereinstimmende Ergebnisse sind in weiteren Simulationsstudien zu finden (z. B. Cook et al. (2009); Friede und Schmidli (2010b)). Im Folgenden liegt das Interesse nun in der Fallzahlanpassung und in der resultierenden Power bei Fehleinschätzung der Störparameter. Zunächst werden die Ergebnisse unter den Bedingungen $\lambda_0^* = \lambda_0$ und $\varphi^* = \varphi$ dargestellt (Tabelle 4.5).

Tabelle 4.5: Charakteristika der ML-Lump und ML-EM Prozedur, wobei λ_0 die Kontrollrate, φ den Shapeparameter, $1 - \beta$ die gewünschte Power und m_0 die initiale Fallzahl widerspiegelt. Weitere Simulationsmerkmale sind in Tabelle 4.4 dargestellt. Die Abkürzung MW steht für Mittelwert, SD für Standardabweichung und Q für Quantil.

λ_0	φ	$1 - \beta$	$m_{0(req)}$	ML-Lump Prozedur						ML-EM Prozedur					
				Power		Fallzahlverteilung				Power		Fallzahlverteilung			
				MW	SD	5Q	50Q	95Q		MW	SD	5Q	50Q	95Q	
2	0.6	0.8	225	0.802	229.9	25.6	190	228	274	0.774	213.8	45.1	138	220	282
		0.9	301	0.903	307.7	29.5	261	307	359	0.870	285.6	58	184	298	369
	0.9	0.8	282	0.805	287.0	29.3	242	286	338	0.765	265.3	54.5	168	276	342
		0.9	372	0.900	384.6	33.7	331	383	443	0.870	354.8	69.9	226	369	452
4	0.6	0.8	170	0.796	174.0	20.8	142	173	210	0.760	157.6	36.2	91	163	211
		0.9	227	0.902	232.5	23.6	195	232	274	0.864	212.6	44.8	128	221	275
	0.9	0.8	227	0.806	231.3	24.1	194	229	272	0.763	211.8	42.4	126	220	271
		0.9	303	0.900	309.4	28.1	266	308	358	0.870	285.1	53.8	174	296	357

Zunächst werden die Ergebnisse der ML-Lump Prozedur erläutert. Die mittlere Power ist vergleichbar mit der angenommenen Power von $1 - \beta$, während die durchschnittlichen Fallzahlen nur geringfügig höher sind als die optimal benötigten Fallzahlen. Im Gegensatz dazu stehen die Resultate der ML-EM Prozedur. Die mittlere Power liegt in allen Szenarien um ca. 3 – 4 % unter der gewünschten Power, was in einer im Mittel nicht ausreichenden Fallzahl begründet ist. Werden die Standardabweichungen der Fallzahlverteilungen verglichen, so zeigt sich eine erheblich größere Streuung im ML-EM Ansatz. Weitere Ergebnisse, die zur gleichen Schlussfolgerung führen, sind in Tabelle 4.6 zusammengefasst.

Um die hohe Variabilität der geschätzten Fallzahl der ML-EM Prozedur zu erklären, wird im Folgenden die Verteilung der aus dem EM-Algorithmus resultierenden Schätzern für das erste Szenario aus Tabelle 4.5 untersucht ($\beta_0 = 0.69$ und $\varphi = 0.6$). Abbildung 4.4 zeigt die Verteilung der geschätzten Störparameter und die Verteilung von $\hat{\beta}_1$. Wird die ursprüngliche initiale Fallzahl von $m_0^* = 225$

Tabelle 4.6: Darstellung der Simulationsergebnisse: die mittlere Fallzahl mit korrespondierendem Standardfehler (se) und Power für das Design ohne Fallzahlانpassung (Fixed) und mit verblindeter bzw. unverblindeter Fallzahlانpassung (ML-EM und ML-Unb). Die Simulationen erfolgten für alle Kombinationen aus Tabelle 4.4 mit 5000 Wiederholungen pro Szenario.

Missspezifizierte Kontrollrate λ_0^{MIS}													
λ_0	ϕ	Power (in %)	Design	0.5 λ_0		0.75 λ_0		λ_0		1.25 λ_0		1.5 λ_0	
				(in %)		(in %)		(in %)		(in %)		(in %)	
				\hat{m}_0 (se)	H ₁	\hat{m}_0 (se)	H ₁	\hat{m}_0 (se)	H ₁	\hat{m}_0 (se)	H ₁	\hat{m}_0 (se)	H ₁
2	0.6	80	Fixed	672	91.30	524	86.90	450	80.20	406	74.70	376	73.20
			ML-Unb	448.6 (0.92)	81.65	447.2 (1.10)	82.45	447.7 (1.18)	81.15	449.7 (1.24)	80.30	448.1 (1.28)	81.35
			ML-EM	426.0 (1.87)	78.30	429.6 (1.97)	75.85	424.6 (2.00)	76.95	431.4 (2.02)	77.05	430.6 (2.10)	77.10
2	0.6	90	Fixed	898	98.10	700	93.25	602	90.09	542	87.25	504	84.25
			ML-Unb	599.8 (1.1)	91.54	600.7 (1.3)	90.03	598.1 (1.3)	91.20	600.0 (1.4)	92.35	599.2 (1.5)	92.00
			ML-EM	577.2 (2.4)	88.80	571.3 (2.4)	86.15	569.6 (2.6)	88.35	574.2 (2.65)	88.20	572 (2.6)	87.60
2	0.9	80	Fixed	784	91.05	638	84.70	564	79.60	520	76.80	490	75.95
			ML-Unb	562.2 (1.2)	80.50	563.7 (1.2)	80.15	563.5 (1.4)	81.95	562.4 (1.4)	80.35	563.8 (1.5)	82.20
			ML-EM	525.6 (2.4)	77.05	531.8 (2.4)	77.60	530.1 (2.5)	77.45	535 (2.5)	76.25	534.7 (2.5)	78.15
2	0.9	90	Fixed	1050	97.00	854	93.30	754	90.05	696	87.25	656	85.30
			ML-Unb	752.4 (1.3)	91.80	753.0 (1.4)	90.15	756.8 (1.5)	91.00	751.9 (1.6)	91.20	751.8 (1.5)	91.75
			ML-EM	700.0 (2.9)	88.60	718.7 (3.0)	87.95	717.3 (3.1)	88.10	712.9 (3.2)	86.70	717.0 (3.2)	86.70
4	0.6	80	Fixed	450	91.15	376	84.15	340	79.90	318	77.10	302	75.80
			ML-Unb	336.2 (0.8)	82.70	336.9 (0.9)	80.75	336.9 (0.92)	80.02	337.3 (0.9)	80.05	336.6 (1.0)	80.25
			ML-EM	315.8 (1.5)	78.75	314.7 (1.6)	76.80	313.3 (1.6)	75.30	313.2 (1.6)	75.35	312.3 (1.7)	76.85
4	0.6	90	Fixed	602	96.80	504	92.80	454	90.00	424	87.70	404	86.15
			ML-Unb	452.5 (0.9)	90.60	450.7 (1.0)	90.95	450.6 (1.1)	91.65	450.9 (1.1)	90.90	450.9 (1.1)	90,75
			ML-EM	425.9 (1.9)	87.50	426.5 (2.0)	86.75	424.7 (2.0)	87.15	424.7 (2.1)	86.50	424.7 (2.0)	88.00
4	0.9	80	Fixed	564	88.30	490	83.30	454	79.75	430	78.30	416	76.25
			ML-Unb	452.1 (0.9)	80.95	452.5 (1.0)	79.55	450.9 (1.1)	81.55	451.8 (1.1)	81.95	449.9 (1.1)	81.70
			ML-EM	427.3 (1.9)	77.95	422.5 (1.9)	74.50	423.6 (2.0)	75.20	424.9 (2.0)	77.35	421.1 (2.0)	76.25
4	0.9	90	Fixed	754	95.55	656	92.35	606	89.60	576	87.90	556	87.25
			ML-Unb	603.7 (1.1)	89.80	605.5 (1.2)	90.60	603.9 (1.2)	90.60	602.5 (1.2)	90.00	602.9 (1.3)	90.75
			ML-EM	573.4 (2.3)	87.75	569.4 (2.4)	87.30	574.2 (2.7)	87.55	566.2 (2.5)	86.70	571.4 (2.5)	87.15

betrachtet, so ist eine bimodale Verteilung von $\hat{\beta}_0$ erkennbar (vgl. erste Zeile in Abbildung 4.4). Somit wird die Kontrollrate in vielen Fällen überschätzt, was zu einer nicht ausreichenden Fallzahl für die gewünschte Power von $1 - \beta$ führt. Die bimodale Verteilung und die starke Streuung des Shape-Parameters führen zu der erheblich größeren Standardabweichung der Fallzahl der ML-EM Prozedur. Des Weiteren zeigt die Verteilung von $\hat{\beta}_1$, dass der EM-Algorithmus in der Hälfte aller Simulationswiederholungen die Kollektive nicht trennen kann ($\hat{\beta}_1 \approx 0$). Erst bei enormen Fallzahlen deutet der EM-Algorithmus auf eine korrekte Funktionsweise hin. In Abbildung 4.4 verändert sich die bimodale Verteilung von $\hat{\beta}_0$ zu einer unimodalen, wobei sich die ursprüngliche Fallzahl verzehn- bzw. verhundertfacht (vgl. zweite bzw. dritte Zeile in Abbildung 4.4). Ähnliche Verbesserungen werden durch größere Effektgrößen erzielt. Da diese Effektstärken und enorme Fallzahlen im klinischen Kontext selten auftreten, scheint die verblindete Fallzahlanpassungsprozedur basierend auf dem EM-Algorithmus auch für Zähldaten nicht angemessen zu sein.

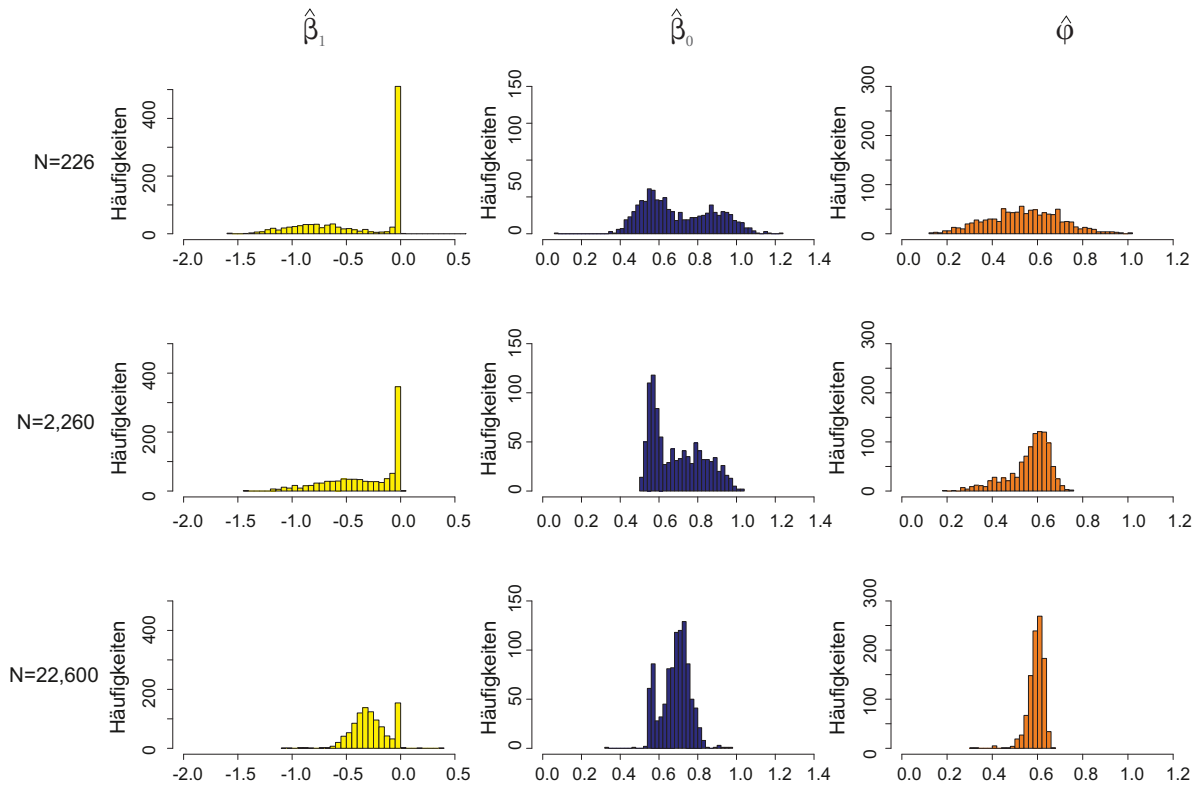


Abbildung 4.4: Die Häufigkeiten von 1000 simulierten Effektschätzern $\hat{\beta}_1$ (linke Spalte), logarithmierten Kontrollratenschätzern $\hat{\beta}_0$ (mittlere Spalte) und Shape-Parameterschätzern $\hat{\phi}$ (rechte Spalte) abhängig von der totalen Fallzahl N (Zeilen) bei Anwendung der ML-EM Prozedur.

Um nun die Ergebnisse der Parameterschätzer in Abhängigkeit von den Konvergenzkriterien zu verdeutlichen, werden die Szenarien aus Tabelle 4.5 abermals aufgegriffen und überprüft. Als Startwert dient der Parametervektor θ_0 aus Gl. 4.10. Die Ergebnisse der Simulationsstudie sind in Ab-

bildung 4.5 für die verschiedenen Konvergenzkriterien dargestellt. Ähnlich wie bei den Resultaten für normalverteilten Daten wird auch hier die starke Abhängigkeit der Ergebnisse vom Konvergenzkriterium beobachtet. Der Algorithmus konvergiert erst bei einem Kriterium, das mindestens 10^{-5} unterschreitet. Wird ein schwächeres Konvergenzkriterium gewählt, so kann es zu Fehleinschätzungen über die Funktionsweise der Prozedur kommen. Ist die Konvergenz jedoch gesichert, so kann eine geringere erwartete Fallzahl mit größerer Standardabweichung beobachtet werden (Abb. 4.5).

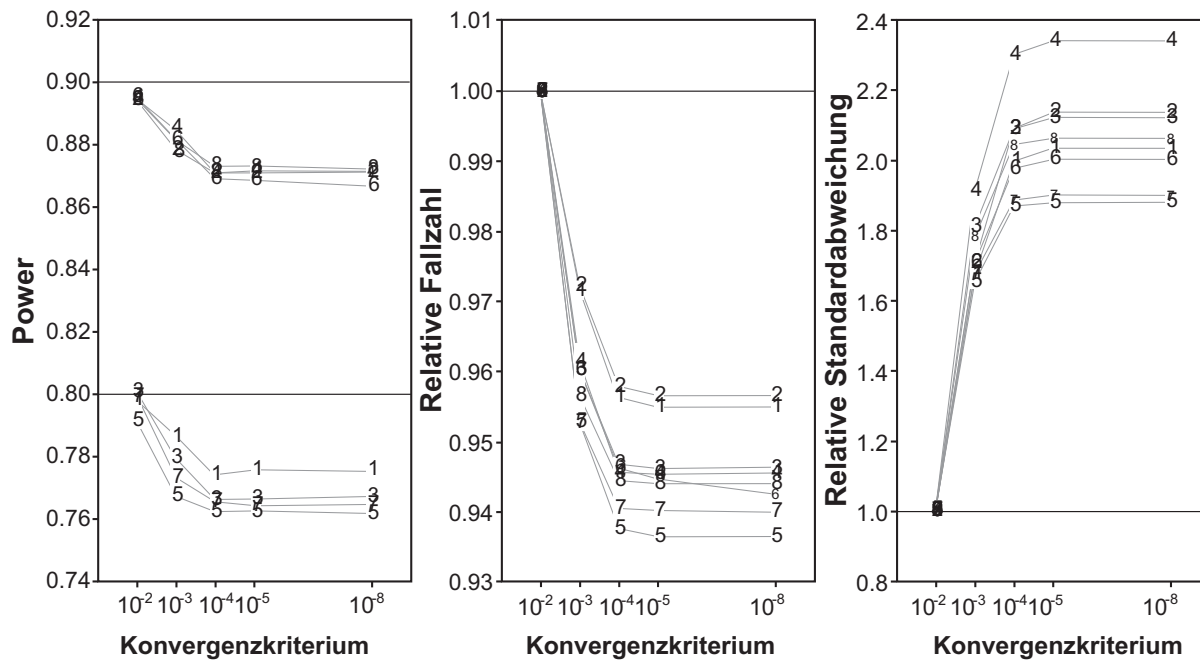


Abbildung 4.5: Simulierte Power (linke Grafik), simulierte relative Fallzahl (mittlere Grafik) und simulierte relative Standardabweichung (rechte Grafik) für die verschiedenen Szenarien aus Tabelle 4.5 abhängig vom Konvergenzkriterium δ . Für die relative Fallzahl und relative Standardabweichung wird das Ergebnis für $\delta = 10^{-2}$ auf 1 gesetzt. Für die stärkeren Konvergenzkriterien wird das Verhältnis berechnet.

Zusammenfassend wird festgestellt, dass die von Cook et al. (2009) beschriebene ML-EM Prozedur eine Anzahl an Defiziten aufweist. Die Parameterschätzung ist, ähnlich wie die Resultate von Friede und Kieser (2002) zeigen, abhängig von der Wahl der Konvergenzkriterien. Konvergiert der EM-Algorithmus nicht, so kann es zu einer Fehlinterpretation der Schätzer kommen. Die Hauptcharakteristika zeigen eine nicht ausreichende Fallzahl für eine vordefinierte Powergröße. Erst bei klinisch irrelevanten Fallzahlen wird eine verbesserte Arbeitsweise beobachtet.

Um zu klären, warum das Verfahren nach Cook et al. (2009) ausreichend gute Resultate zeigt, wurde zusätzlich ein unverblindetes Verfahren, welches in Abschnitt 3.3.2 vorgestellt wurde, in die Simu-

lationsstudie einbezogen. Es ist zu erkennen, dass der unverblindete Ansatz (ML-Unb) zu ähnlich guten Ergebnissen führt wie die ML-Lump Prozedur (vgl. Tabelle 4.5 und 4.6). Vergleichbare Resultate wurden bereits für das unverblindete Verfahren im Poisson-Modell in Absatz 4.1.2 vorgestellt. Infolgedessen liegt die Vermutung nahe, dass die Resultate der unverblindeten und verblindeten Prozedur in Cook et al. (2009) verwechselt wurden.

4.2.4 Beispiel

Zur Demonstration der aufgezeigten Problematiken der ML-EM Prozedur wird im Folgenden auf Daten der „Asymptomatic Cardiac Ischemia Pilot“ (ACIP) Studie (Rogers et al., 1995; Davies et al., 1997) zurückgegriffen, die vom US National Heart, Lung and Blood Institute (NHLBI) gesponsert wurde. Das Ziel der randomisierten, kontrollierten klinischen Studie ist die Erforschung der Effizienz und Sicherheit dreier Ansätze zur Vorbeugung asymptomatischer Myokardischämien. Des Weiteren wurden die Effekte auf weitere bedeutende kardiale Ereignisse wie beispielsweise Herzinfarkt oder Tod untersucht. Der primäre Endpunkt der Studie war die Eliminierung der Ischämien im 48-Stunden Elektrokardiogramm nach 12 Wochen Behandlungszeit. Da kein Zugriff auf Patientendaten möglich ist, müssen der unverblindete Schätzer der Kontrollrate $\hat{\lambda}_0 = 1.146$, der Schätzer der Effektstärke $\exp(\hat{\beta}_1) = 0.452$ und der Schätzer des Shape-Parameters $\hat{\phi} = 1.986$ zur Generierung eines Beispieldatensatzes benutzt werden (vgl. Cook et al. (2009)). Die resultierenden Schätzer des generierten Datensatzes, der in Tabelle 4.7 (i) dargestellt ist, unterscheiden sich maximal um ein Hundertstel von den genannten Schätzern. Für den EM-Algorithmus werden der angenommene Effekt von $\beta_{1(0)} = \log(0.7)$ und die logarithmierte mittlere Rate aller Ereignisse $\beta_{0(0)}$ als Startwerte verwendet. Der Startwert für den Shape-Parameter wird über ein Negativ-Binomial-Regressionsmodell ohne Berücksichtigung der Gruppenzugehörigkeit ermittelt. Zusätzlich wird der angenommene Effekt von $\beta_A = \log(0.7)$ im Fallzahlenanpassungsschritt benutzt. Abbildung 4.6 (i) zeigt die geschätzte

Tabelle 4.7: Häufigkeiten auftretender kardialen Ischämien für zwei generierte Szenarien, wobei in (i) der wahre Effekt den Wert $\exp(\beta_1) = 0.452$ und in (ii) $\exp(\beta_1) = 0.7$ annimmt.

		(i) $e^{\beta_1} = 0.452$	(ii) $e^{\beta_1} = 0.7$
Anzahl an Ereignissen	ECG-geleitete Therapie (N=170)	Angina-geleitete Therapie (N=170)	Angina-geleitete Therapie (N=170)
0	89 (49 %)	120 (71 %)	107 (63 %)
1	43 (25 %)	33 (25 %)	25 (15 %)
2	16 (9 %)	9 (14 %)	23 (14 %)
3	7 (4 %)	3 (2 %)	5 (3 %)
4	3 (2 %)	1 (1 %)	4 (2 %)
≥ 5	12 (7 %)	4 (2 %)	6 (4 %)

Fallzahl in der Zwischenauswertung abhängig vom Konvergenzkriterium δ . Analog zu den Simulationsergebnissen weist der Schätzer der Kontrollrate eine bimodale Verteilung bei Konvergenz des Algorithmus auf. Somit wird die Kontrollrate in vielen Szenarien überschätzt, was zu einer Reduzierung der Fallzahl bei Konvergenz führt. Eine Aufrechterhaltung der Studienpower ist demzufolge

nicht mehr möglich. Im Vergleich zur unverblindeten Fallzahlanpassungsprozedur führt der verblindete Ansatz zu einem verzerrten Schätzer von λ_0 , da die wahre Effektgröße nicht der angenommenen Effektgröße entspricht. Die Verzerrung kann durch den Ausdruck $\hat{\lambda}_0 = 2\lambda/(1 + \exp(\beta_1^*)) = 0.97$ angegeben werden. Daraus resultiert eine im Vergleich zum unverblindeten Ansatz um 17 % erhöhte Fallzahl. Wie zuvor beschrieben und an anderen Beispielen diskutiert, sollten in derartigen Situationen andere adaptive Designs in Erwägung gezogen werden (vgl. Abschnitt 4.1.3 bzw. 4.1.4). Um eine bessere Vergleichbarkeit zu erlangen, wird der gleiche Simulationsaufbau mit der Annahme $(\exp(\beta_1) = \exp(\beta_1^*) = 0.7)$ überprüft. Ein weiterer Datensatz ist in Tabelle 4.7 (ii) dargestellt. Die Bedingungen der Startwerte und Konvergenzkriterien an den EM-Algorithmus sind unverändert. Abbildung 4.6 (ii) zeigt die Simulationsergebnisse. Liegt ein schwaches Konvergenzkriterium (z. B. 10^{-2}) vor, so bricht der Algorithmus nach wenigen Iterationen - also nahe den Startwerten - ab. Dieses Verhalten führt erneut zu einer geeigneten geschätzten Fallzahl. Im Gegensatz dazu resultiert bei gesicherter Konvergenz die bimodale Verteilung des Schätzers der Kontrollrate in einer ungenügenden Fallzahlschätzung. Für die ML-EM Prozedur ist der Schätzer der Kontrollrate $\hat{\lambda}_0 = 1.47$ im Vergleich zu $\hat{\lambda}_0 = 1.15$ aus dem ML-Lump Ansatz im Mittel überschätzt. Auch in diesem Szenario ist eine Aufrechterhaltung der Studienpower nicht mehr möglich.

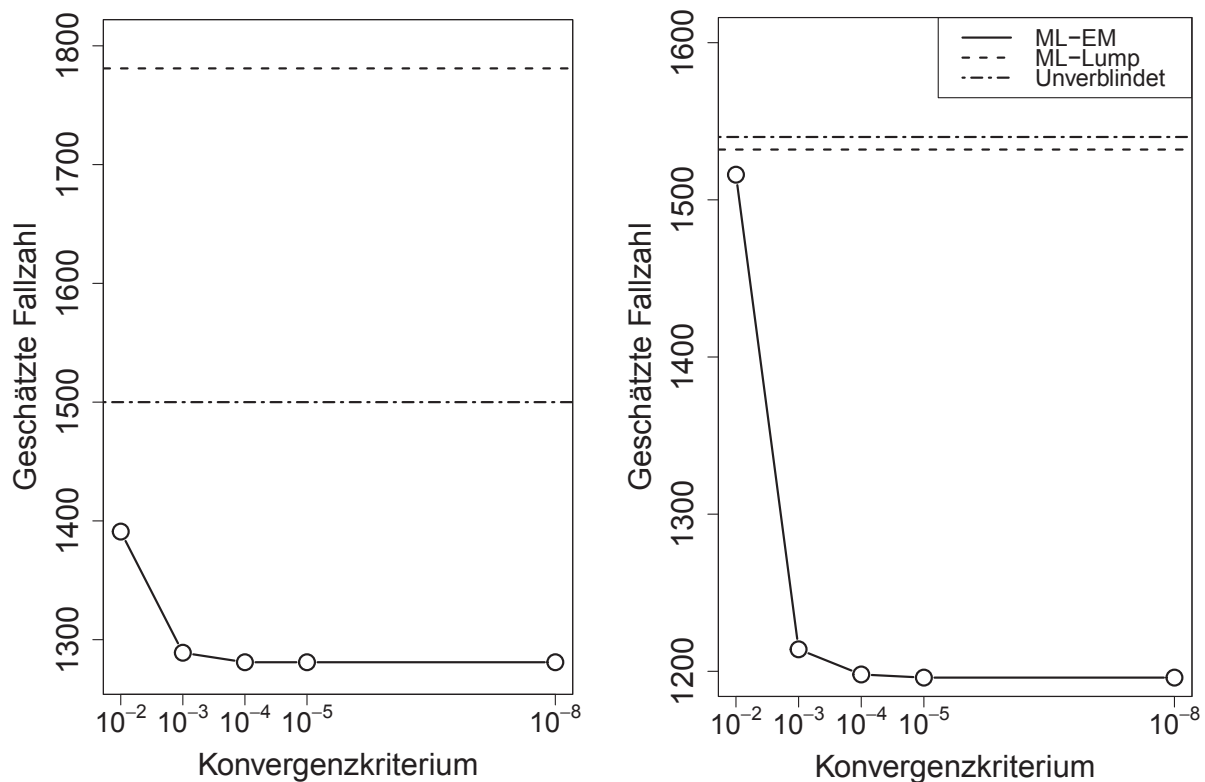


Abbildung 4.6: Darstellung der Fallzahlen durch Anwendung der ML-EM Prozedur in Abhängigkeit des Konvergenzkriteriums für die Szenarien der Tabelle 4.3 ((i):=linke und (ii):= rechte Graphik). Die horizontalen Linien repräsentieren die Resultate der unverblindeten Prozedur und ML-Lump Prozedur.

4.2.5 Robustheitsanalyse

Die zuvor betrachteten Fallzahlenanpassungsprozeduren basieren auf dem Negativ-Binomial-Modell in Gleichung 2.7, wobei der Shape-Parameter φ in beiden Behandlungsgruppen identisch ist. Im folgenden Abschnitt wird die Sensitivität dieser Annahme untersucht. Dazu werden zwei Shape-Parameter φ_0 und φ_1 für die Kontroll- bzw. Behandlungsgruppe eingeführt. In der nachfolgenden Simulationsstudie wird zwischen zwei Szenarien unterschieden. Der Shape-Parameter der Kontrollrate φ_0 bleibt bei variierendem φ_1 konstant und umgekehrt. Die vorgestellten Szenarien sind in Tabelle 4.8 zusammengefasst.

Tabelle 4.8: Simulationsaufbau der zwei Szenarien, wobei λ_0 die Kontrollrate, φ_1 den konstanten Shape-Parameter und φ_0 den variablen Parameter der Kontrollgruppe darstellen (1). In (2) ist die Beziehung von φ_0 und φ_1 vertauscht.

Abbildung	λ_0	m_0	nsim	Signifikanzniveau α	φ_0	φ_1	$1 - \beta$
(1)	2 (○)	225	10000	0.05 (zweiseitig)	0.1, 0.2, ..., 3	0.6	0.8
	4 (□)	170					
(2)	2 (○)	225	10000	0.05 (zweiseitig)	0.6	0.1, 0.2, ..., 3	0.8
	4 (□)	170					

Die ML- und QL-Lump Prozeduren sind in die Untersuchung eingeschlossen, da eine weitere Analyse der EM-Algorithmus-basierenden Fallzahlenanpassungsstrategie angesichts der aufgetretenen Defizite abwegig wäre. Der Quasi-Likelihood Ansatz bietet insbesondere bei Unsicherheiten hinsichtlich der Entstehung der Heterogenität gewisse Robustheitseigenschaften (McCullagh und Nelder, 1989) und wird aus diesem Grund in den Fokus gerückt. Die initiale Fallzahl wird über die Annahme $\varphi_0 = \varphi_1 = 0.6$ berechnet. Die Ergebnisse der Studie sind in Abbildung 4.7 zusammengefasst. Beide Prozeduren sind bis zu einem gewissen Maße robust gegenüber der Modellabweichung. Nur bei sehr großen Differenzen zwischen den Shape-Parametern (1) : $|\varphi_0 - \varphi_{0a}|$ (z. B. $\varphi_0 = 3$ und $\varphi_1 = 0.6$) oder (2) : $|\varphi_1 - \varphi_{1a}|$ (z. B. $\varphi_0 = 0.6$ und $\varphi_1 = 3$) wird die gewünschte Power um 3 % Punkte verfehlt (vgl. Abb. 4.7a). Aus der Abbildung 4.7b geht hervor, dass die QL-Methode (grau) in einigen Szenarien eine höhere mittlere Fallzahl benötigt als die ML-Lump Methode (schwarz), um eine ähnliche Power zu erreichen. Aus der Abbildung 4.7c geht hervor, dass ein enormer Unterschied in den simulierten Streuungen der Fallzahlverteilungen zwischen den beiden Prozeduren existiert. Die Schätzung der Dispersion im QL-Ansatz folgt mittels Momentenmethode, die in einer größeren Streuung im Vergleich zum Maximum-Likelihood Ansatz resultiert.

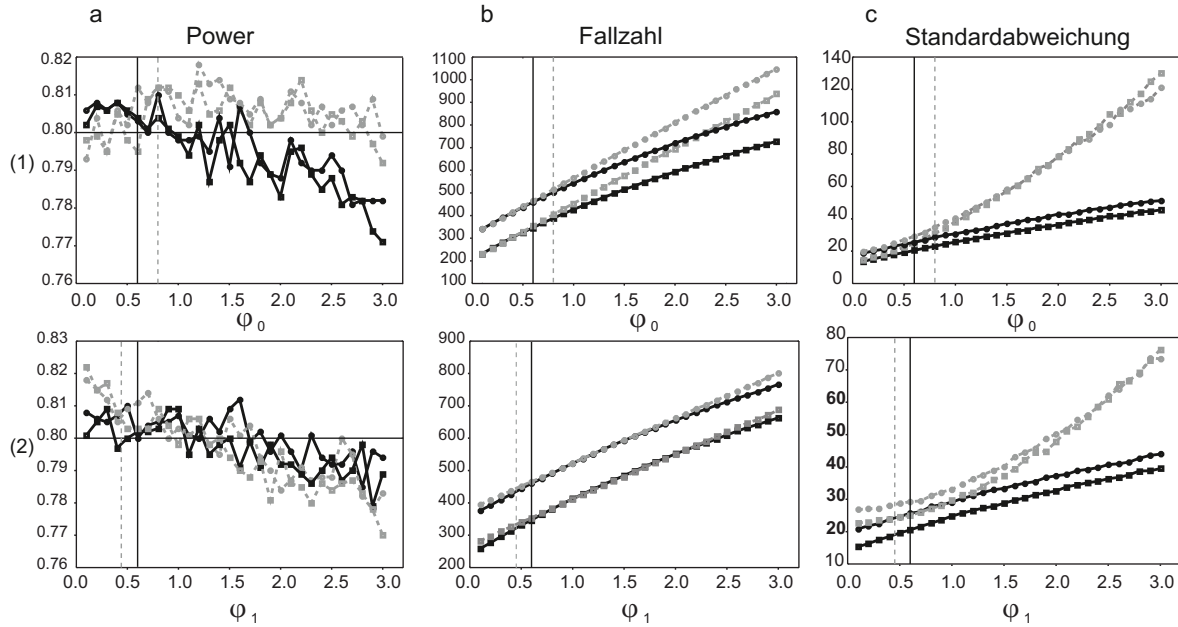


Abbildung 4.7: Darstellung der Ergebnisse der Simulationsstudie aus Tabelle 4.8. In der linken Grafik ist die Power, in der mittleren die durchschnittliche Fallzahl und rechts die zugehörige Standardabweichung der QL-Lump (grau) und der ML-Lump (schwarz) Prozeduren abhängig vom Shape-Parameter φ_1 , φ_0 aufgetragen. Die vertikalen Linien spiegeln die Werte wider, für die einerseits die Bedingung des Quasi-Likelihood Ansatzes oder aber die des Negativ-Binomial-Modells $\varphi_0 = \varphi_1 = 0.6$ korrekt ist.

Zusammenfassend bieten beide Methoden gewisse Robustheitseigenschaften, falls die Modellannahmen verletzt werden. Die beobachtete kleinere Streuung der Fallzahl führt jedoch dazu, dass die Anwendung der Maximum-Likelihood Methode präferiert werden sollte.

KAPITEL 5

ERWEITERUNG DER PROZEDUR FÜR ZEITVARIIERENDE RATEN

5.1 Zählprozesse

Poisson-verteilte Zufallsvariablen können - anders als zuvor erläutert (vgl. Abschnitt 2.1.1) - als Zuwachs eines Poisson-Prozesses gesehen werden. In den folgenden Abschnitten wird die Modellierung rekurrenter Ereignisse aus Sicht der Zähl- und Intensitätsprozesse beschrieben, die sich im weiteren Verlauf als besonders nützlich erweisen. Zugunsten einer besseren Lesbarkeit wird an dieser Stelle auf eine Einführung in die Standarddefinition für stochastische Prozesse verzichtet - nähere Erläuterungen sind Anhang A zu entnehmen.

5.1.1 Notation und mathematische Grundlagen

Sei $N(s, t)$ die Anzahl an Ereignissen, die im Zeitraum $(s, t]$ für ein Individuum beobachtet werden. Es wird davon ausgegangen, dass der Beobachtungszeitraum zum Zeitpunkt $s = 0$ beginnt und keine Ereignisse zu diesem Punkt vorliegen ($N(0) = 0$). Das Auftreten der Ereignisse wird durch den definierten Zählprozess $\{N(t) = N(0, t), t \geq 0\}$ beschrieben (vgl. A.7).

Eine generelle Beschreibung des Modells wird durch die Wahrscheinlichkeitsverteilung der auftretenden Ereignisse in einem schmalen Intervall $[t, t + \Delta t]$ gegeben der Geschichte $H(t)$ des Zählprozesses zum Zeitpunkt t bereitgestellt. $H(t)$ ergibt sich durch

$$H(t) = \{N(s) : 0 \leq s < t\} \text{ mit } t > 0,$$

wobei die Anzahl an Ereignissen im Zeitintervall $[t, t + \Delta t]$ mit $\Delta N(t) = N(t + \Delta t) - N(t)$ bezeichnet wird.

Bevor die Wahrscheinlichkeitsfunktion der auftretenden Ereignisse definiert wird, wird die

Definition der *Intensitätsfunktion* eingeführt:

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N(t) = 1|H(t))}{\Delta t}. \quad (5.1)$$

Sie beschreibt die (sofortige) Wahrscheinlichkeit von dem Auftreten eines Ereignisses zum Zeitpunkt t , bedingt auf die Geschichte des Prozesses. Im stetigen Fall wird davon ausgegangen, dass höchstens ein Ereignis zu einem Zeitpunkt auftreten kann. Unter Verwendung der Definition der Intensitätsfunktion wird schließlich eine generelle Beschreibung der Modelle in folgendem Satz gegeben.

Satz 5.1 (vgl. Cook und Lawless (2007)) *Gegeben die Geschichte $H(r)$ ist die Dichtefunktion für das Auftreten von genau $n \geq 0$ Ereignissen an den Zeitpunkten $t_1 < \dots < t_n$ für einen Prozess mit Intensitätsfunktion (Gl. 5.1) in einem Intervall $[r, s]$ definiert durch:*

$$\prod_{j=1}^n \lambda(t_j|H(t_j)) \cdot \exp\left(-\int_r^s \lambda(u|H(u)) du\right). \quad (5.2)$$

BEWEIS. Für die Berechnung der Wahrscheinlichkeitsfunktion für das Auftreten der Ereignisse werden Resultate der Produktintegration verwendet. Eine genaue Herleitung des Satzes ist durch Cook und Lawless (2007) gegeben. \square

Auf Basis dieser Definitionen und Sätze wird der Poisson-Prozess im folgenden Abschnitt eingeführt.

5.1.2 Poisson-Prozess

Ein spezieller stochastischer Prozess (siehe A.4) ist der nach dem französischen Physiker und Mathematiker Siméon Denis Poisson benannte *Poisson-Prozess*. In der vorliegenden Arbeit wird der Poisson-Prozess mit Hilfe der im vorherigen Abschnitt definierten Intensitätsfunktion eingeführt. Für den Poisson-Prozess hat die Intensitätsfunktion aus Gl. 5.1 die Form $\lambda(t|H(t)) = \lambda(t)$, wobei $\lambda(t)$ eine nichtnegative integrierbare Funktion und nicht von der Geschichte des Prozesses abhängig ist. An die kumulative Intensitätsfunktion $\Lambda(t) = \int_0^t \lambda(u) du$ wird die Annahme der Endlichkeit für alle $t > 0$ gestellt. Zunächst wird zwischen einem homogenen und inhomogenen Poisson-Prozess unterschieden. Ein homogener Poisson-Prozess liegt bei einer konstanten Intensitätsfunktion $\lambda(t) = \lambda$ vor und ist somit als wichtiger Spezialfall des inhomogenen Prozesses zu verstehen. In homogenen Poisson-Prozessen wird von stationären Zuwächsen gesprochen. Weitere Eigenschaften gelten für den (in)homogenen Poisson-Prozess.

Lemma 5.1

1. $N(s, t)$ ist poissonverteilt mit Parameter $\Lambda(s, t) = \Lambda(t) - \Lambda(s)$ für $0 \leq s < t$ (für den homogenen Prozess folgt: $\Lambda(s, t) = \lambda(t - s)$).
2. Für nicht-überlappende Zeitintervalle $(s, t]$ und $(s_1, t_1]$ sind die Zufallsvariablen $N(s, t)$ und $N(s_1, t_1)$ unabhängig.

BEWEIS. Die erste Eigenschaft folgt unmittelbar mit dem Satz 5.1, der die Dichtefunktion für das Eintreten von genau n Ereignissen an den Zeitpunkten $t_1 < \dots < t_n$ im Intervall $(s, t]$ wie folgt gibt (Cook und Lawless, 2007):

$$\prod_{j=1}^n \lambda(t_j) \cdot \exp(-\Lambda(s, t)).$$

Für die Randwahrscheinlichkeit gilt dann:

$$P(n \text{ Ereignisse} \in (s, t]) = \left[\int \dots \int \left[\prod_{j=1}^n \lambda(t_j) \right] dt_1 \dots dt_n \right] \cdot \exp(-\Lambda(s, t)), \quad (5.3)$$

wobei das multiple Integral über den Bereich $s < t_1 < \dots < t_n \leq t$ geht. Mit der Symmetrie des Integranden in t_1, \dots, t_n folgt für den Ausdruck in Gl. 5.3:

$$\begin{aligned} P(n \text{ Ereignisse} \in (s, t]) &= \left[\int \dots \int \left[\prod_{j=1}^n \lambda(t_j) \right] dt_1 \dots dt_n \right] \cdot \exp(-\Lambda(s, t)) \\ &= \left[(n!)^{-1} \int_s^t \dots \int_s^t \prod_{j=1}^n \lambda(t_j) dt_1 \dots dt_n \right] \cdot \exp(-\Lambda(s, t)) \\ &= \left[(n!)^{-1} \prod_{j=1}^n \left[\int_s^t \lambda(t_j) dt_j \right] \right] \cdot \exp(-\Lambda(s, t)) \\ &= (n!)^{-1} \Lambda(s, t)^n \cdot \exp(-\Lambda(s, t)) \end{aligned}$$

Das Resultat zeigt die Wahrscheinlichkeitsfunktion der Poisson-Verteilung. Die zweite Eigenschaft folgt mit $t < s_1$ und dem Merkmal, dass die Zufallsvariable $N(s_1, t_1)$ unabhängig von ihrer Geschichte ist. Damit muss sie auch von $N(s, t)$ unabhängig sein. \square

Liegt der Spezialfall eines homogenen Poisson-Prozesses vor, so folgen die Zähldaten schließlich einer Poisson-Verteilung mit konstanter Rate λ (vgl. Abschnitt 2.1.1 ff.).

Im folgenden Abschnitt wird der nach Friede und Schmidli (2010b) vorgeschlagene Ansatz zur Fallzahlanpassung auf zeitvariierende Baseline-Raten $\lambda_0(t)$ erweitert. Dafür werden - abweichend zu vorherigen Kapiteln - Zufallsvariablen N_{ij} betrachtet, die einem inhomogenen Poisson-Prozess folgen.

5.2 Statistisches Modell, Hypothesentest und Fallzahlplanung

5.2.1 Statistisches Modell

Die unabhängigen Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, mit den Realisierungen n_{ij} , die an den Zeitpunkten $t_{ij1} < \dots < t_{ijy_{ij}}$ über ein Beobachtungsintervall von $(0, T_{ij}]$ auftreten, folgen einem inhomogenen Poisson-Prozess mit zugehöriger Intensitätsfunktion $\lambda_i(t, \boldsymbol{\theta})$. Im Folgenden wird von einer parametrischen Baseline-Intensitätsfunktion ausgegangen, die von dem Parametervektor

θ abhängt. Dann ist das statistische Modell wie folgt gegeben:

$$N_{ij}(T_{ij}) \sim \text{Pois}(\Lambda_i(T_{ij}, \theta)) \quad i = 0, 1, \quad j = 1, \dots, m_i,$$

wobei $\Lambda_i(u, \theta) = \int_0^u \lambda_i(x, \theta) dx$ darstellt (vgl. Lemma 5.1). Das gegebene statistische Modell wird durch Einführung eines zufälligen Effekts auf ein Gamma-Cox-Modell erweitert. Die bedingte Intensitätsfunktion für den Zählprozess $\{N_{ij}(t)|U_{ij}\}$ ist durch

$$\lambda_i(t, \theta|U_{ij} = u_{ij}) = u_{ij} \lambda_i(t, \theta) \quad i = 0, 1, \quad j = 1, \dots, m_i,$$

gegeben, wobei $U_{ij} \sim \Gamma\left(\frac{1}{\varphi}, \varphi\right)$ ist (Schneider et al., 2013b). Somit hat das statistische Modell die folgende Form:

$$N_{ij}(T_{ij})|U_{ij} = u_{ij} \sim \text{Pois}(u_{ij} \Lambda_i(T_{ij}, \theta)).$$

Analog zu den vorherigen Abschnitten wird im nächsten Absatz ein Regressionsmodell vorgestellt, das eine Modellierung der erwarteten Häufigkeiten in Abhängigkeit der fixen Kovariablen x_i ermöglicht.

5.2.2 Regressionsmodell

Seien die Zufallsvariablen N_{ij} , $i = 0, 1$, $j = 1, \dots, m_i$, Poisson-verteilt mit der Rate $u_{ij} \Lambda_i(t, \theta)$. Das Interesse liegt im Einfluss der Behandlung, welches in dem folgenden Regressionsmodell durch die Kovariable x_i ausgedrückt wird:

$$E(N_{ij}(T_{ij})|U_{ij} = u_{ij}) = u_{ij} \Lambda_0(T_{ij}, \theta) \exp(x_i \beta_1) \quad i = 0, 1, \quad j = 1, \dots, m_i, \quad (5.4)$$

wobei $\Lambda_0(T_{ij}, \theta) = \int_0^u \lambda_0(x, \theta) dx$ und $\lambda_0(t, \theta)$ die Baseline-Intensitätsfunktion widerspiegelt. Das proportionale Intensitätsmodell bekommt durch den vorhandenen multiplikativen Effekt $\exp(x_i \beta_1)$ mit $\beta_1 = \log(\lambda_1 / \lambda_0)$ seinen Namen.

5.2.3 Schätzer, Hypothesentest und Fallzahlplanung

Zunächst wird in diesem Abschnitt die Dichtefunktion für das Auftreten von genau n_{ij} Realisierungen an den Zeitpunkten $t_{ij1} < \dots < t_{ijn_{ij}}$ über ein Beobachtungsintervall von $(0, T_{ij}]$ für Patient j in Behandlungsgruppe i mit Hilfe von Satz 5.1 zu

$$L_{ijk}(\varphi, \beta_1, \theta | x_i, n_{ij}, t_{ijk}) = \int_0^\infty \prod_{k=1}^{n_{ij}} u_{ij} \lambda_i(t_{ijk}, \theta, \beta_1) \cdot \exp\left(-\int_0^{T_{ij}} u_{ij} \lambda_i(v, \theta, \beta_1) dv\right) f(u_{ij}) du_{ij}$$

berechnet, wobei $f(\cdot)$ die Wahrscheinlichkeitsfunktion der Gamma-Verteilung darstellt. Eine marginale Darstellung der Likelihood für alle $j = 1, \dots, m_i$ Patienten in Behandlungsgruppe $i = 0, 1$ mit den Realisierungen n_{ij} an den Zeitpunkten $t_{ij1} < \dots < t_{ijn_{ij}}$ über ein Beobachtungsintervall von $(0, T_{ij}]$ ist demzufolge durch das Produkt im Ausdruck

$$L(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} \left(\prod_{k=1}^{n_{ij}} \frac{\lambda_0(t_{ijk}, \boldsymbol{\theta})}{\Lambda_0(T_{ij}, \boldsymbol{\theta})} \right) \times \frac{\Gamma(\varphi^{-1} + n_{ij})}{\Gamma(\varphi^{-1}) n_{ij}!} \frac{(\varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(x_i \beta_1))^{n_{ij}}}{(1 + \varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(x_i \beta_1))^{n_{ij} + \varphi^{-1}}} \quad (5.5)$$

gegeben (z. B. Lawless (1987b, 1995); Cook und Lawless (2002)). Wird der Fokus auf den zweiten Faktor gelegt, so ist zu beobachten, dass er nur von der Summe der Realisierungen pro Individuum abhängt und zum anderen die Wahrscheinlichkeitsfunktion der Negativ-Binomial-Verteilung widerspiegelt. Die Likelihood in Gl. 5.5 wird als vollständige Likelihood bezeichnet. Die Ereigniszeitpunkte beeinflussen lediglich den ersten Term. Diese Erkenntnisse werden im Folgenden für den Hypothesentest und die Fallzahlplanung verwendet.

Der zugrunde liegende Intensitätsprozess $\lambda(t, \boldsymbol{\theta})$ ist oftmals schwer zu charakterisieren. Aus diesem Grund werden zunächst die aggregierten Schätzer mit den resultierenden Schätzern aus einer vollständigen Likelihood Analyse verglichen. Betrachtet wird die folgende Log-Likelihood:

$$l(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk}) = \sum_{i=0}^1 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} \log \left(\frac{\lambda_0(t_{ijk}, \boldsymbol{\theta})}{\Lambda_0(T_{ij}, \boldsymbol{\theta})} \right) + \underbrace{\sum_{i=0}^1 \sum_{j=1}^{m_i} \log \left(\frac{\Gamma(\varphi^{-1} + n_{ij})}{\Gamma(\varphi^{-1}) n_{ij}!} \frac{(\varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(x_i \beta_1))^{n_{ij}}}{(1 + \varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(x_i \beta_1))^{n_{ij} + \varphi^{-1}}} \right)}_{\log(L_{NB}(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk}))}. \quad (5.6)$$

Die partiellen Ableitungen von $l(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk})$ nach β_1 und φ sind identisch mit den Ableitungen im Negativ-Binomial-Regressionsmodell in Gl. 2.10 und Gl. 2.11 und werden mit g_{β_1} und g_{φ} bezeichnet. Die partielle Ableitung von $l(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk})$ nach $\boldsymbol{\theta}$ für die vollständige Daten Log-Likelihood ist durch

$$g_{\boldsymbol{\theta}_f} = \frac{\partial l(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk})}{\partial \boldsymbol{\theta}} = \sum_{i=0}^1 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log \left(\frac{\lambda_0(t_{ijk}, \boldsymbol{\theta})}{\Lambda_0(T_{ij}, \boldsymbol{\theta})} \right) \right] + \underbrace{\frac{\partial \log(L_{NB}(\varphi, \beta_1, \boldsymbol{\theta} | x_i, n_{ij}, t_{ijk}))}{\partial \boldsymbol{\theta}}}_{g_{\boldsymbol{\theta}_{ag}}}. \quad (5.7)$$

gegeben. Das bringt zum Ausdruck, dass eine Analyse basierend auf den aggregierten Daten, auf den partiellen Ableitungen nach β_1 und φ und dem zweiten Summanden der Log-Likelihood Gleichung in Gl. 5.7 beruht. Im Folgenden bezeichne $\boldsymbol{\kappa}_f = (\hat{\beta}_1, \hat{\varphi}, \hat{\boldsymbol{\theta}})$ den Schätzer resultierend aus den Lösungen des Gleichungssystems $\mathbf{g}_f = (g_{\beta_1}, g_{\varphi}, g_{\boldsymbol{\theta}_f})' = \mathbf{0}'$ und bezeichne $\boldsymbol{\kappa}_{ag} = (\tilde{\beta}_1, \tilde{\varphi}, \tilde{\boldsymbol{\theta}}_{ag})$ den Schätzer aus Lösung des Systems $\mathbf{g}_{ag} = (g_{\beta_1}, g_{\varphi}, g_{\boldsymbol{\theta}_{ag}})' = \mathbf{0}'$. Die asymptotische Normalität der Schätzer $\boldsymbol{\kappa}_f$ und $\boldsymbol{\kappa}_{ag}$ und die dazugehörigen Kovarianzmatrizen werden von Dean und Balshaw (1997) im Appendix B gegeben.

Zur Durchführung von Hypothesentest und Fallzahlplanung wird die Varianz von $\hat{\beta}_1$ benötigt. Diese hängt jedoch von der Form der Intensitätsfunktion und folglich von dem Parametervektor $\boldsymbol{\theta}$ ab. Ei-

ne Spezifikation dieser Größen ist im klinischen Kontext nahezu unmöglich. Infolgedessen wird auf Theorem 2 nach Dean und Balshaw (1997) zurückgegriffen, welches einen anderen Ansatz ermöglicht. Für die Annahme gleicher Beobachtungszeiträume aller Patienten $T_{ij} = T$ oder mindestens balancierter Nachbeobachtungszeiten in beiden Behandlungsgruppen $\sum_{j=1}^{n_0} T_{0j} \approx \sum_{j=1}^{n_1} T_{1j}$ weisen die Effektschätzer $\tilde{\beta}_1$ aus den aggregierten Daten und $\hat{\beta}_1$ aus der vollständigen Likelihood eine sehr hohe asymptotische relative Effizienz auf. Besonders für kleine Shape-Parameter konvergiert die asymptotische Effizienz gegen 1. Eine Simulationsstudie nach Dean und Balshaw (1997) spiegelt diese Eigenschaft in klinisch relevanten Situationen wider. In den klinischen Studien in multipler Sklerose ist die Annahme über die Nachbeobachtungszeiten nicht abwegig, solange kein systematischer Drop-Out vorliegt, da dies das derzeit gängige Studiendesign ist.

Das Interesse liegt, wie in Abschnitt 2.2.3 bereits erläutert, in der Untersuchung der Hypothese $H_0 : \beta_1 \geq 0$ gegen die Alternative $H_1 : \beta_1 < 0$. Im Folgenden wird der Schätzer $\tilde{\beta}_1$ zum Testen herangezogen und die korrespondierende Waldtyp Statistik T_{ag} ist im Ausdruck

$$T_{ag} = \frac{\tilde{\beta}_1}{\sqrt{\left(\frac{1}{\sum_{j=1}^{m_{0j}} n_{0j}} + \frac{1}{\sum_{j=1}^{m_{1j}} n_{1j}} + \hat{\varphi} \left(\frac{m_{0j} + m_{1j}}{m_{0j} m_{1j}} \right) \right)}} \stackrel{\mathcal{D}}{\sim} N(0, 1)$$

gegeben (vgl. Abschnitt 2.2.3). Für den Anwender impliziert diese Feststellung, dass am Ende einer klinischen Studie ein Negativ-Binomial-Modell für konstante Raten zum Schätzen des Effektes β_1 angepasst werden kann. Daraus folgend kann die Fallzahlplanung für einen angenommenen Effekt β_1^* , zur angenommenen mittleren Kontrollrate $\Lambda(T, \theta)^*$ und für eine gewünschten Power $1 - \beta$ mit Hilfe der Fallzahlformel für die Kontrollgruppe

$$m_0 = \frac{(z_{\alpha/2} + z_{1-\beta})^2}{\beta_1^{*2}} \left(\frac{1}{\Lambda_0(T, \theta)^*} \left(1 + \frac{1}{k \exp(\beta_1^*)} \right) + \varphi^* (1 + 1/k) \right) \quad (5.8)$$

durchgeführt werden. Die Herleitung folgt analog zur Fallzahlformel in Gleichung 2.15 für konstante Raten (Keene et al., 2007; Friede und Schmidli, 2010b). Durch das Einsetzen einer konstanten Baseline-Intensitätsfunktion $\lambda_0(t) = \lambda_0$ in Gl. 2.15 ist die Äquivalenz der beiden Ausdrücke durch $\Lambda_0(T, \theta)^* = \lambda_0^* T$ zu sehen. Insbesondere soll durch die Notation $\Lambda_0(T, \theta)^*$ deutlich gemacht werden, dass keine Annahme an den Parametervektor θ sondern nur an die mittlere Rate $\Lambda_0(T, \theta)$ über den Zeitraum T gemacht werden muss.

5.3 Verblindete Fallzahlplanung bei zeitvariierenden Zählraten (BSSR-TREND)

Analog zu Abschnitt 3.3 gibt es zwei Möglichkeiten eine verblindete Fallzahlplanung im proportionalen Intensitätsmodell aus Abschnitt 5.2 durchzuführen. Notationen und Ideen werden aus Abschnitt 3.2 übernommen. Da die Nachbeobachtungszeiten im Data-Review zwischen den Patienten variieren und eine aggregierte Analyse somit zu ineffizienten Schätzern für den Parametervektor θ

führt, muss in der Zwischenauswertung die vollständige Likelihood betrachtet werden (Dean und Balshaw, 1997; Schneider et al., 2013b). Um die Kontrollrate aus den verblindeten Daten zu schätzen, kann die folgende Mischverteilung mit festem Behandlungseffekt β_1^* herangezogen werden.

$$L(\varphi, \boldsymbol{\theta} | \beta_1^*, n_{ij}, t_{ijk}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} \left[\frac{1}{k+1} \left(\prod_{k=1}^{n_{ij}} \frac{\lambda_0(t_{ijk}, \boldsymbol{\theta})}{\Lambda_0(T_{ij}, \boldsymbol{\theta})} \right) \times \frac{\Gamma(\varphi^{-1} + n_{ij})}{\Gamma(\varphi^{-1}) n_{ij}!} \frac{(\varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}))^{n_{ij}}}{(1 + \varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}))^{n_{ij} + \varphi^{-1}}} \right. \\ \left. + \frac{k}{k+1} \left(\prod_{k=1}^{n_{ij}} \frac{\lambda_0(t_{ijk}, \boldsymbol{\theta})}{\Lambda_0(T_{ij}, \boldsymbol{\theta})} \right) \times \frac{\Gamma(\varphi^{-1} + n_{ij})}{\Gamma(\varphi^{-1}) n_{ij}!} \frac{(\varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(\beta_1^*))^{n_{ij}}}{(1 + \varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(\beta_1^*))^{n_{ij} + \varphi^{-1}}} \right]$$

Die Likelihood wird nun hinsichtlich des Störvektors $\boldsymbol{\theta}$ und des Shape-Parameters φ mit Hilfe numerischer Routinen maximiert. Die asymptotisch erwartungstreuen Schätzer liefern in einer Fallzahlanpassung nicht die gewünschte Power am Ende der klinischen Studie. Im Falle der Normalverteilung und ebenso im konstanten Modell wurde ein ähnliches Verhalten in vorherigen Kapiteln diskutiert. Aus diesem Grund wird der nach Friede und Schmidli (2010a,b) beschriebene Ansatz zur Fallzahlanpassung auf das statistische Modell übertragen. Hierzu wird auf die Likelihood aus Gl. 5.5 zurückgegriffen

$$L(\varphi, \boldsymbol{\theta}_b | n_{ij}, t_{ijk}) = \prod_{i=0}^1 \prod_{j=1}^{m_i} \left(\prod_{k=1}^{n_{ij}} \frac{\lambda_0(t_{ijk}, \boldsymbol{\theta}_b)}{\Lambda_0(T_{ij}, \boldsymbol{\theta}_b)} \right) \times \frac{\Gamma(\varphi^{-1} + n_{ij})}{\Gamma(\varphi^{-1}) n_{ij}!} \frac{(\varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}_b))^{n_{ij}}}{(1 + \varphi \Lambda_0(T_{ij}, \boldsymbol{\theta}_b))^{n_{ij} + \varphi^{-1}}},$$

wobei die Behandlungszugehörigkeit für alle Patienten vernachlässigt wird (d.h. $x_i = 0$ für $i = 0, 1$). Die zugehörigen Schätzer von $\boldsymbol{\theta}_b$ und φ besitzen keine geschlossene Lösung, so dass erneut numerische Verfahren zur Optimierung der Log-Likelihood benutzt werden. Zudem wird durch das Ignorieren der Gruppenzugehörigkeit nicht der Parametervektor $\boldsymbol{\theta}$ geschätzt sondern ein verblindeter Vektor $\hat{\boldsymbol{\theta}}_b$. Analog zu dem Modell mit konstanten Raten kann die resultierende Rate $\Lambda(T, \hat{\boldsymbol{\theta}}_b)$ als mittlere Rate gedeutet werden. Da der Behandlungseffekt multiplikativ in das Modell eingeht, ist es möglich, einen verblindeten Kontrollratenschätzer durch Adjustierung der mittleren Rate mit einer Effektannahme β_1^* durch den Ausdruck

$$\Lambda_0(T, \hat{\boldsymbol{\theta}}_b) = \frac{k+1}{1 + k \exp(\beta_1^*)} \Lambda(T, \hat{\boldsymbol{\theta}}_b)$$

zu erhalten. Der verblindete Kontrollratenschätzer wird in die Fallzahlformel 5.8 eingesetzt. Die mittlere geschätzte Rate führt durch die Gleichung

$$\hat{m}_0 = \frac{(z_{\alpha/2} + z_{1-\beta})^2}{\beta_1^{*2}} \left(\frac{1}{\Lambda(T, \hat{\boldsymbol{\theta}}_b)} \left(\frac{(1 + k \exp(\beta_1^*))^2}{(k+1) k \exp(\beta_1^*)} \right) + \hat{\varphi} (1 + 1/k) \right)$$

zu einem äquivalenten Ergebnis.

5.4 Zeittrends in klinischen Studien im Krankheitsbild der schubförmigen multiplen Sklerose

Um die im vorherigen Kapitel eingeführte Fallzahlenanpassungsstrategie zu untersuchen, wird eine log-lineare Baseline Intensitätsfunktion gewählt:

$$\lambda_0(t, \boldsymbol{\theta}) = \exp(\alpha_0 + \alpha_1 t). \quad (5.9)$$

Die Funktion ist abhängig von der Zeit t und dem Parametervektor $\boldsymbol{\theta} = (\alpha_0, \alpha_1)'$. Der Vorteil einer log-linearen Baseline-Funktion liegt darin, dass für alle Parametermöglichkeiten $\boldsymbol{\theta} = (\alpha_0, \alpha_1)' \in \mathbb{R}^2$ die resultierende Rate positiv ist. Welche Werte für den Achsenabschnitt und die Steigung geeignet sind, wird im nächsten Abschnitt erläutert. Zunächst wird eine Gamma-verteilte Zufallsvariable U_{ij} eingeführt, die jedem Patienten eine individuelle Rate zuweist. Die Intensitätsfunktion hat somit folgende Form:

$$\lambda_i(t, \alpha_0, \alpha_1 | U_{ij} = u_{ij}) = u_{ij} \lambda_0(t, \alpha_0, \alpha_1) \exp(\beta_1 x_i)$$

mit der entsprechenden kumulativen Intensitätsfunktion

$$\Lambda_i(t, \alpha_0, \alpha_1 | U_{ij} = u_{ij}) = u_{ij} \frac{1}{\alpha_1} \exp(\alpha_0) (\exp(\alpha_1 t) - 1) \exp(\beta_1 x_i). \quad (5.10)$$

Der Effekt Λ_1/Λ_0 ist unabhängig von der Wahl der Intensitätsfunktion $\lambda_0(t, \boldsymbol{\theta})$; somit wird das Modell als proportionales Intensitätsmodell bezeichnet. Das Einsetzen von Gl. 5.10 in die Likelihood in Gl. 5.5 liefert eine Log-Likelihood der Form

$$\begin{aligned} l(\alpha_0, \alpha_1, \varphi, \beta_1) = & \sum_{i=0}^1 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} [\alpha_0 + \alpha_1 t_{ijk}] + \sum_{i=0}^1 \sum_{j=1}^{m_i} n_{ij} (\log(\varphi) + x_i \beta_1) \\ & + \sum_{i=0}^1 \sum_{j=1}^{m_i} (\log(\Gamma(n_{ij} + 1/\varphi)) - \log(\Gamma(1/\varphi))) \\ & - \sum_{i=0}^1 \sum_{j=1}^{m_i} [(n_{ij} + \varphi^{-1}) \log(1 + \varphi / \alpha_1 \exp(\alpha_0) \exp(x_i \beta_1) (\exp(\alpha_1 T_{ij}) - 1))]. \end{aligned}$$

Das folgende Gleichungssystem

$$\left(\frac{\partial l(\alpha_0, \alpha_1, \varphi, \beta_1)}{\partial \beta_1}, \frac{\partial l(\alpha_0, \alpha_1, \varphi, \beta_1)}{\partial \varphi}, \frac{\partial l(\alpha_0, \alpha_1, \varphi, \beta_1)}{\partial \alpha_0}, \frac{\partial l(\alpha_0, \alpha_1, \varphi, \beta_1)}{\partial \alpha_1} \right)' = \mathbf{0}$$

muss numerisch gelöst werden und resultiert in einem asymptotisch normalverteilten Parameterschätzer $\boldsymbol{\kappa}_f = (\hat{\beta}_1, \hat{\varphi}, \hat{\alpha}_0, \hat{\alpha}_1)$. Die Verteilung der Schätzer folgt mit den Herleitungen und Verweisen in Abschnitt 5.2.3.

Für die initiale Fallzahlplanung einer klinischen Studie wird die Fallzahlformel in Gl. 5.8 unter Einbeziehung der kumulativen Kontrollrate herangezogen. Somit kann die bekannte Formel abhängig von dem Parametervektor $\boldsymbol{\theta}$ beschrieben werden, der aus praktischer Sicht jedoch schwer zu interpretieren und klassifizieren ist. Mit den vorgestellten Ergebnissen aus Abschnitt 5.2.3 reicht es aus

eine mittlere Rate $\Lambda_0(T, \boldsymbol{\theta})^*$ zu spezifizieren, da der Effektschätzer aus der vollständigen Likelihood-Analyse $\hat{\beta}_1$ eine asymptotisch hohe relative Effizienz zum aggregierten Schätzer $\tilde{\beta}_1$ aufweist. Lediglich im Data-Review wird die verblindete Log-Likelihood

$$\begin{aligned} l(\alpha, \alpha_1, \varphi) = & \sum_{i=0}^1 \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} [\alpha + \alpha_1 t_{ijk}] + \sum_{i=0}^1 \sum_{j=1}^{m_i} n_{ij} \log(\varphi) \\ & + \sum_{i=0}^1 \sum_{j=1}^{m_i} (\log(\Gamma(n_{ij} + 1/\varphi)) - \log(\Gamma(1/\varphi))) \\ & - \sum_{i=0}^1 \sum_{j=1}^{m_i} [(n_{ij} + \varphi^{-1}) \log(1 + \varphi 1/\alpha_1 \exp(\alpha) (\exp(\alpha_1 T_{ij}) - 1))]. \end{aligned} \quad (5.11)$$

benötigt. Es wird festgestellt, dass die Ignoranz des multiplikativen Effektes $\exp(\beta_1)$ ausschließlich einen Einfluss auf den Achsenabschnitt α_0 der log-linearen Funktion hat. Infolgedessen wird mit Hilfe der verblindeten Likelihood ein Schätzer für den Gesamtachsenabschnitt α , der im folgenden Ausdruck gegeben ist, bestimmt:

$$\exp(\alpha) = \frac{\exp(\alpha_0) + k \exp(\alpha_0) \exp(\beta_1)}{k + 1} = \exp(\alpha_0) \left(\frac{1 + k \exp(\beta_1)}{k + 1} \right). \quad (5.12)$$

Die resultierenden Schätzer $\hat{\alpha}, \hat{\alpha}_1$ werden im Fallzahlanpassungsschritt zur Bestimmung der gemeinsamen kumulativen Intensitätsfunktion $\Lambda(T, \hat{\alpha}, \hat{\alpha}_1)$ benötigt. Ein verblindeter Kontrollratenschätzer wird durch Adjustierung mit dem angenommenen Effekt durch den Faktor $\frac{k+1}{1+k \exp(\beta_1)}$ berechnet (Gleichung 5.12). Dieser Schätzer ist durch den Ausdruck

$$\Lambda_0(T, \hat{\boldsymbol{\theta}}_b) = \frac{k + 1}{1 + k \exp(\beta_1^*)} \Lambda(T, \hat{\boldsymbol{\theta}}_b)$$

gegeben. Werden die beiden Schätzer $\Lambda_0(T, \hat{\boldsymbol{\theta}}_b)$ und $\hat{\varphi}_b$ in die Fallzahlformel (Gl. 5.8) eingesetzt, so resultiert eine aktualisierte Fallzahl \hat{m}_0 zu einer gewünschten Power von $1 - \beta$ und einem angenommenen Effekt β_1^* . Die finale Testprozedur kann bei nicht zu großen Abweichungen von der Annahme $\sum_{j=1}^{n_0} T_{0j} \approx \sum_{j=1}^{n_1} T_{1j}$ mit den aggregierten Daten durchgeführt werden (Dean und Balshaw, 1997).

5.5 Simulationsstudie motiviert durch Studien in der multiplen Sklerose

In diesem Abschnitt wird das erweiterte Verfahren zur Fallzahlanpassung mit zeitvariierenden Zähl-daten in einer ausführlichen Simulationsstudie hinsichtlich Power, der Fehlerkontrolle und der Fallzahlverteilung untersucht. Um eine eventuelle Überlegenheit und Unterschiede zwischen der neuen BSSR-TREND Prozedur und dem bekannten Ansatz für konstante Raten nach Friede und Schmidli (2010b) zu erkennen, wird dieses Verfahren in die Simulationsstudie eingeschlossen und im weiteren Verlauf als BSSR-CONST bezeichnet. Für die folgende Simulationsstudie liegt die log-lineare Baseline-Intensitätsfunktion $\lambda_0(t, \alpha_0, \alpha_1) = \exp(\alpha_0 + t\alpha_1)$ aus Abschnitt 5.4 zugrunde, so dass sich das

bedingte Regressionsmodell zu

$$N_{ij}(T_{ij})|U_{ij} = u_{ij} \sim \text{Pois}(u_{ij}\Lambda_i(T_{ij}, \boldsymbol{\theta}, \beta_1)),$$

$$E(N_{ij}(T_{ij})|U_{ij} = u_{ij}) = u_{ij}\Lambda_0(T_{ij}, \boldsymbol{\theta}) \exp(x_i\beta_1)$$

mit $\Lambda_i(t, \alpha_0, \alpha_1|U_{ij} = u_{ij}) = u_{ij} \frac{1}{\alpha_1} \exp(\alpha_0) (\exp(\alpha_1 t) - 1) \exp(\beta_1 x_i)$ ergibt. Zunächst stellt sich die Frage, welche Werte von α_0 und α_1 den Schubratenprozess der Patienten angemessen beschreiben. Für die Bestimmung adäquater Werte wird auf eine Meta-Analyse in schubförmiger multipler Sklerose nach Nicholas et al. (2012) zurückgegriffen. In der Meta-Analyse wurden die Schubraten von sieben randomisierten kontrollierten Studien des ersten Jahres den angegebenen Schubraten des zweiten Jahres gegenübergestellt (Abb. 1.2). Der Quotient der Jahresraten (Schubrate Jahr 1/Schubrate Jahr 2) der verschiedenen Studien wurde zu einem Gesamteffekt zusammengefasst und sollte bei „keinem Zeiteffekt“ nicht signifikant von der 1 abweichen. Die Ergebnisse zeigen jedoch zwei signifikante Verhältnisse der 1-12 vs. 13-24 Monatsraten von 1.25 (1.1 – 1.43) für die Placebo und 1.4 (1.19 – 1.63) für die Behandlungsgruppe (Konfidenzintervall). Das Verhältnis der 1-12 vs. 13-24 Raten wird nun in das parametrische Modell übertragen. Es gilt

$$\psi = \frac{\Lambda_0(T/2, \boldsymbol{\theta}) - \Lambda_0(0, \boldsymbol{\theta})}{\Lambda_0(T, \boldsymbol{\theta}) - \Lambda_0(T/2, \boldsymbol{\theta})},$$

wobei die Wahl von $T = 2$ das geschilderte Szenario widerspiegelt. Es ist erkennbar, dass nach der Substitution von $\Lambda_0(t, \boldsymbol{\theta})$ durch Gl. 5.10 der Quotient unabhängig von der Wahl des Achsenabschnittes ist:

$$\psi = \frac{\exp(\alpha_1) - 1}{\exp(2\alpha_1) - \exp(\alpha_1)}. \quad (5.13)$$

Zur Bestimmung der Steigung α_1 wird der maximal beobachtete Quotient der Placebo- und Behandlungsgruppe in Formel 5.13 eingesetzt. Für $\psi = 1.63$ ergibt sich die Lösung $\alpha_1 = \log(1/\psi) = -0.5$. Demzufolge deckt die Steigung, korrespondierend zu den Werten nach Nicholas et al. (2012), im log-linearen Modell einen Bereich von $\alpha_1 = 0$ bis $\alpha_1 = -0.5$ ab. Ein ähnliches Ergebnis wird erhalten, wenn die Resultate der 1-6 vs. 7-12 Monatsraten eingesetzt werden. Geeignete Werte für den Achsenabschnitt α_0 werden durch die Arbeit von Steinvorth et al. (2013) gewonnen, die eine Kontrollrate von $\lambda_0 = \exp(\alpha_0) \in [0.4 - 0.6]$ liefert.

In den folgenden Simulationsstudien wird die Studiendauer auf $T = 1$ oder $T = 2$ Jahre gesetzt. Die Rekrutierungsperiode wird im Folgenden mit T_{rp} bezeichnet und in den Variationen $T_{rp} = 1, 1.5$ oder $T_{rp} = 2$ untersucht. Die Zwischenauswertung wird am Ende der Rekrutierungszeit, wenn alle ursprünglich geplanten Patienten eingeschlossen sind, durchgeführt. Somit kann es zu keiner Reduzierung der initial geplanten Fallzahl kommen, jedoch sind andererseits genügend Informationen über die Patienten zur Schätzung erhältlich. Für zwei Szenarien $T = 2$ und $T_{rp} = 2$ bzw. $T = 2$ und $T_{rp} = 1$ ist das Rekrutierungsschema beispielhaft in Abbildung 5.1 dargestellt. Eine längere Rekrutierungszeit T_{rp} im Verhältnis zu der Studiendauer führt zu zunehmend vollständigeren Patienteninformationen (z. B. $T = 2$ und $T_{rp} = 2$, Abb. 5.1a). Andererseits ist durch das Szenario $T = 2$ und

$T_{rp} = 1$ (Abb. 5.1b) ersichtlich, dass eine zügige Rekrutierung zu einem geringen Informationsstand führt.

Alle folgenden Szenarien wurden mit $n_{\text{sim}} = 10.000$ Wiederholungen durchgeführt. Das Signifikanzniveau ist auf $\alpha = 0.025$ (einseitig), die gewünschte Power auf $1 - \beta = 0.8$ und der angenommene wie auch der wahre Behandlungseffekt auf $\exp(\beta_1^*) = \exp(\beta_1) = 0.75$ gesetzt. Eine genaue Beschreibung der Simulation eines inhomogenen Poisson-Prozesses ist im Anhang B zu finden.

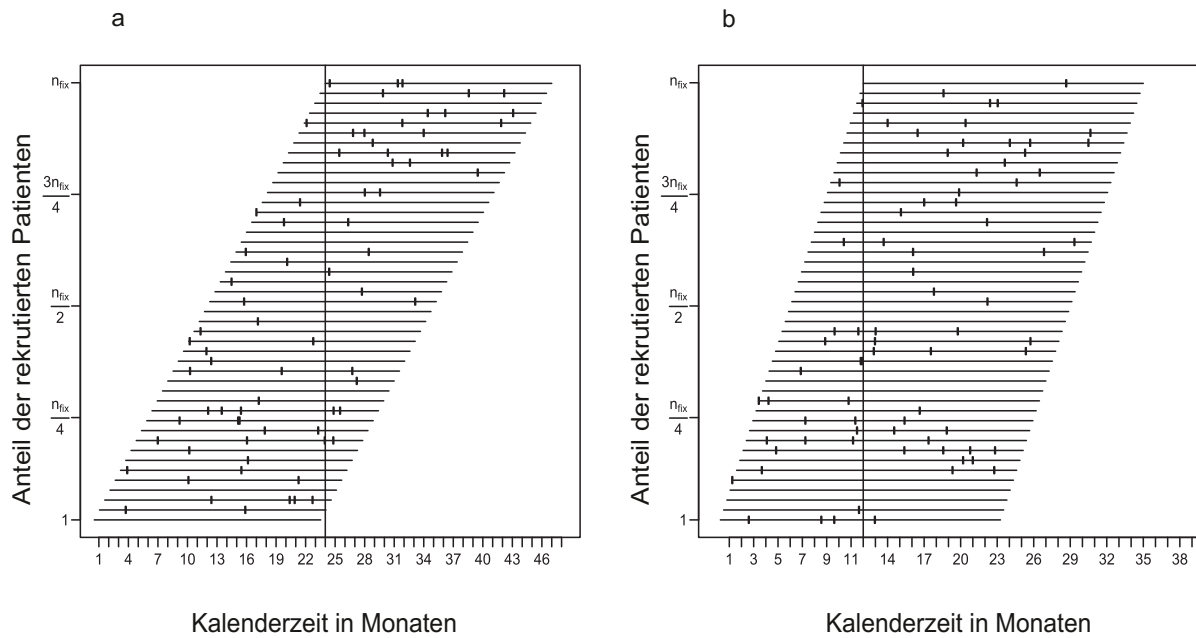


Abbildung 5.1: Beispielhafter Studienablauf für eine 2-Jahres-Studie $T = 2$ mit zweijähriger Rekrutierungsperiode in der linken sowie einjähriger Rekrutierungsperiode in der rechten Abbildung. Die vertikale Linie spiegelt den Zeitpunkt der Zwischenauswertung wider, an dem n_{fix} Patienten rekrutiert sind. Die Striche stellen Ereignisse der Patienten dar.

5.5.1 Power

Zunächst wird das Verhalten der entwickelten verblindeten Fallzahlenanpassungsprozedur BSSR-TREND und der bekannten Fallzahlenanpassungsprozedur BSSR-CONST hinsichtlich der Aufrechterhaltung der gewünschten Studienpower mit Hilfe zweier Szenarien überprüft. Aus Vergleichszwecken werden zusätzlich die Ergebnisse ohne die Anwendung von Fallzahlenanpassungsstrategien („fixed design“) angegeben. Die Szenarien sind wie folgt gegeben:

- Szenario I:

Für das erste Szenario werden eine mittlere Kontrollrate $\Lambda_0^*(2) = 1.2$, ein Shape-Parameter $\varphi^* = \varphi = 0.6$ und kein vorliegender Zeittrend $\alpha_1^* = 0$ angenommen. Dieses Setting führt zu einer initialen Kontrollfallzahl von $m_0 = 298$.

- Szenario II:

Für das zweite Szenario werden eine Kontrollrate $\Lambda_0(2) = 1.2$, ein Shape-Parameter $\varphi^* = 0$ und kein vorliegender Zeittrend $\alpha_1^* = 0$ angenommen, was in einer initialen Fallzahl von $m_0 = 184$ resultiert.

Abbildung 5.2a zeigt die simulierte Power der verblindeten Prozeduren und die des fixen Designs in Abhängigkeit der Steigung α_1 . Zudem wurden verschiedene Rekrutierungsperioden T_{rp} in Betracht gezogen, um mehrere Szenarien der Datengüte widerzuspiegeln. Es ist erkennbar, dass die Anwendung eines fixen Designs nur in Situationen, in denen alle Störparameter richtig spezifiziert sind, zu brauchbaren Ergebnissen führt ($\alpha_1 = 0$). In allen anderen Szenarien wird eine drastisch reduzierte Studienpower beobachtet.

Die BSSR-CONST Prozedur nutzt lediglich die aggregierten Daten zur Schätzung der Störparameter. Ein potentieller Trend in den Daten wird demzufolge nicht aufgedeckt. In Situationen, in denen ein starker Trend ($\alpha_1 = -0.5$) und nur wenige Daten zur Zwischenauswertung vorliegen ($T_{rp} = 1$), wird die Power von 80 % nicht erreicht. Liegen umfangreichere Patienteninformationen vor, so verbessert sich die BSSR-CONST Prozedur (Abb. 5.2a, $T = 2$ und $T_{rp} = 2$). Um die bereits erläuterte Verzerrung von $\hat{\varphi}_b$ zu verdeutlichen, ist in Tabelle 5.1 die Verteilung von $\hat{\varphi}_b$ für verschiedene Steigungen zusammengefasst. Mit steigendem α_1 wird die Verzerrung größer, da die Nachbeobachtungszeiten zur Zwischenauswertung stark variieren, was zu einer größeren Variation in den Raten führt, da sich diese unbeachtet über die Zeit verändern. Diese Verzerrung reicht jedoch nicht aus, um die Fallzahl derart zu erhöhen, dass ein Powerverlust bei starken Zeittrends kompensiert wird.

Die entwickelte BSSR-TREND Prozedur ist entgegen aller beschriebener Ergebnisse robust gegenüber einer irrtümlichen Annahme an einen vorliegenden Trend. In allen Szenarien liefert der Ansatz Ergebnisse, die nahe der gewünschten Power von $1 - \beta = 0.8$ liegen. Liegt kein Trend vor ($\alpha_1 = 0$), so sind alle benötigten Patienten schon eingeschlossen und die Fallzahl kann lediglich unnötig erhöht werden. Das spiegelt sich in einer Power von 85 % wider und ist ebenfalls für die BSSR-CONST Prozedur erkennbar.

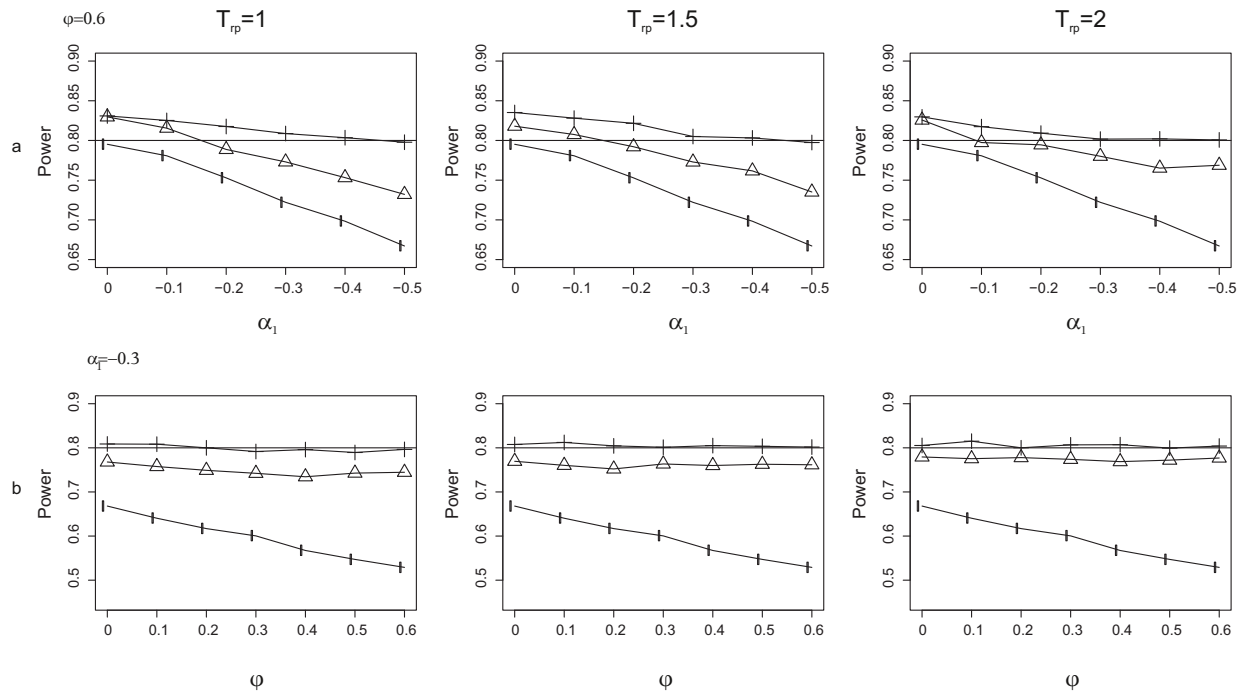
Im Folgenden wird die Auswirkung der Falschannahmen für den Shape-Parameter und für die Steigung der Intensitätsfunktion untersucht. In der Planungsphase der klinischen Studie wird zunächst von keiner interindividuellen Variabilität $\varphi^* = 0$ und keinem Zeittrend $\alpha_1^* = 0$ ausgegangen. Abbildung 5.2b zeigt nun die simulierte Power der Prozeduren in Abhängigkeit des wahren Shape-Parameters φ . Zusätzlich wird ein wahrer Zeittrend, der in der Steigung von $\alpha_1 = -0.3$ ausgedrückt wird, simuliert.

Werden keine Fallzahlanpassungsprozeduren benutzt, muss mit einem drastischen Powerverlust gerechnet werden. In allen Szenarien sind die wahren Störparameterwerte ungleich den angenommen

Störparameterwerten, so dass ein fixes Design keine adäquaten Ergebnisse zeigen kann. In Situationen, in denen eine starke interindividuelle Variabilität vorliegt, ist dieser Verlust mit etwa 30 % anzugeben.

Die BSSR-CONST Prozedur ist robust gegen die irrtümliche Annahme an den Shape-Parameter φ^* , kann den vorhandenen Zeittrend jedoch nicht kompensieren. Das sich die Prozedur mit günstigerer Datenlage verbessert, wurde in Szenario I beschrieben und ist ebenfalls hier erkennbar.

Nur die entwickelte BSSR-TREND Prozedur liefert robuste Ergebnisse, die nahe der gewünschten Power von $1 - \beta = 0.8$ liegen. Zusammenfassend wird beobachtet, dass die neue Anpassungsstrategie die erwarteten Eigenschaften hinsichtlich der Aufrechterhaltung der gewünschten Studienpower besitzt. Es existieren Situationen, in denen die BSSR-CONST bzw. das fixe Design zu adäquaten Ergebnissen führen. Das trifft zu, wenn ein positiver Trend in den Daten zu finden ist ($\alpha_1 > 0$). Dennoch wäre in solchen Szenarien eine zu hohe Power der Studien wahrscheinlich. Da die Richtung des Zeittrends nicht immer bekannt ist, sollte das erweiterte Verfahren BSSR-TREND priorisiert werden.



Abbildungung 5.2: Power der verblindeten Prozeduren BSSR-TREND (+) und BSSR-CONST (Δ) abhängig von der Steigung α_1 (erste Zeile, $\varphi = 0.6$) und vom Shape-Parameter φ (untere Zeile, $\alpha_1 = -0.3$) für verschiedene Rekrutierungsperioden T_{rp} . Zu Vergleichszwecken sind die Ergebnisse für ein Design ohne Fallzahlanpassung angegeben (|).

Tabelle 5.1: Resultierende Verteilung von $\hat{\phi}$ bei der Anwendung der BSSR-CONST Prozedur, wobei das Verhalten bei verschiedenen Steigungen α_1 untersucht wird. Die Studiendauer wird mit $T = 2$, die Rekrutierungsperiode mit $T_{rp} = 1.5$, $\exp(\alpha_0) = 0.6$, $\phi = 0.6$ und der Effekt mit $\exp(\beta_1) = 0.75$ festgelegt.

α_1	Mittelwert	(SD)	(Min - Max)
0	0.628	(0.172)	(0.357-0.921)
-0.1	0.626	(0.180)	(0.344-0.931)
-0.2	0.632	(0.187)	(0.341-0.948)
-0.3	0.635	(0.199)	(0.321-0.976)
-0.4	0.640	(0.210)	(0.309-1.000)
-0.5	0.643	(0.224)	(0.294-1.019)

5.5.2 Fehler 1. Art und Fallzahlverteilung

Um die Fehlerkontrolle und Fallzahlverteilung der entwickelten BSSR-TREND Prozedur zu analysieren, wird eine weitere Simulationsstudie mit einer aussagekräftigen Anzahl an Szenarien durchgeführt (Tab. 5.2). Es werden alle 48 Kombinationen der Einflussfaktoren simuliert, wobei alle Störparameter bis auf die Steigung α_1 , die in der Planungsphase durch $\alpha_1^* = 0$ spezifiziert wird, als richtig angenommen werden. In der Planungsphase einer klinischen Studie ergibt sich somit eine mittlere Kontrollrate von $\Lambda_0(T) = 0.4T$ oder $\Lambda_0(T) = 0.6T$ für $T = 1, 2$. Die übrigen Simulationsgrundlagen können Abschnitt 5.5.1 entnommen werden.

Tabelle 5.3 zeigt die zusammengefassten Ergebnisse der 48 simulierten Szenarien. Um die Wahrscheinlichkeit einen Fehler 1. Art zu begehen zu charakterisieren, werden die Szenarien unter der Nullhypothese simuliert. Es kann keine Inflation des nominalen Niveaus α beobachtet werden, da alle Resultate im Monte Carlo Fehlerbereich von $[0.044, 0.055]$ liegen (Tab. 5.3). Des Weiteren können die Ergebnisse der Powersimulationen aus Abschnitt 5.5.1 auf die erweiterte Simulationsstudie übertragen werden. Der Mittelwert aller Simulationen liegt bei 80.5 %, somit nahe an der gewünschten Power von $1 - \beta = 0.8$; nur in 5 Szenarien wurde die Power nicht erreicht. Das Minimum von 79.60 % zeigt, dass die ermittelten Abweichungen marginal sind.

Tabelle 5.2: Einflussfaktoren auf die Simulationsstudie

Faktor		Szenario
Studiendauer	T	1, 2
Rekrutierungsperiode	T_{rp}	1, 1.5, 2
Achsenabschnitt	e^{α_0}	0.4, 0.6
Steigung	α_1	-0.5, -0.3
Shape-Parameter	ϕ	0.35, 0.7

Um eine Aussage über die Fallzahlverteilung tätigen zu können, wird der Quotient aus der geschätzten und der tatsächlich benötigten Fallzahl \hat{n}_0/n_{req} sowie der Variationskoeffizient von \hat{n}_0 , der die

relative Standardabweichung $CV = \frac{SD(\hat{n}_0)}{\hat{n}_0}$ beschreibt, in der Analyse betrachtet. Für beide Endpunkte wird ein vollständiger Versuchsplan aufgestellt, der alle möglichen Faktorkombinationen aus Tabelle 5.2 prüft. Die Faktoren werden durch die Faktorstufen $-1, 1$ oder $-1, 0, 1$ beschrieben. Beinahe alle Effekte - bis auf die Interaktion von Studienlänge und Rekrutierungsperiode - sind nahe 0. Die genannte Interaktion der beiden Endpunkte ist in Abbildung 5.3 erkennbar.

Tabelle 5.3: Verfahrenseigenschaften für alle 48 Kombinationen aus Tabelle 5.2. Die Abkürzung CV steht für den Variationskoeffizient und SD für die Standardabweichung.

Eigenschaften	Mittelwert	(SD)	Median	(Min - Max)
Power [%]	80.5	(0.46)	80.5	(79.6 - 81.5)
Fehler 1. Art [%]	5.12	(0.2)	5.16	(4.58 - 5.45)
Variationskoeffizient von \hat{n}_0	0.13	(0.05)	0.11	(0.08 - 0.24)
\hat{n}_0 / n_{req}	1.025	(0.013)	1.020	(1.011 - 1.061)

Die Resultate belegen, dass eine lange Studiendauer in Kombination mit einer kurzen Rekrutierungsperiode (z. B. $T = 2$ and $T_{rp} = 1$) zu einer größeren Unsicherheit in der Schätzung der Fallzahl ausgedrückt durch den Variationskoeffizient (Abb. 5.3b) und zu einer nun durchschnittlich 4.5 % zu hoch geschätzten Fallzahl (Abb. 5.3a) führt. In Situationen, in denen sich die Datenlage verbessert, reduziert sich die Unsicherheit, wovon die BSSR-TREND Prozedur profitiert. Zusammenfassend sollte der Zeitpunkt der Zwischenauswertung idealerweise mit dem Rekrutierungsmuster abgestimmt werden, so dass eventuell vorliegende Trends mit einer möglichst kleinen Variation geschätzt werden können und zu adäquaten Fallzahlen führen.

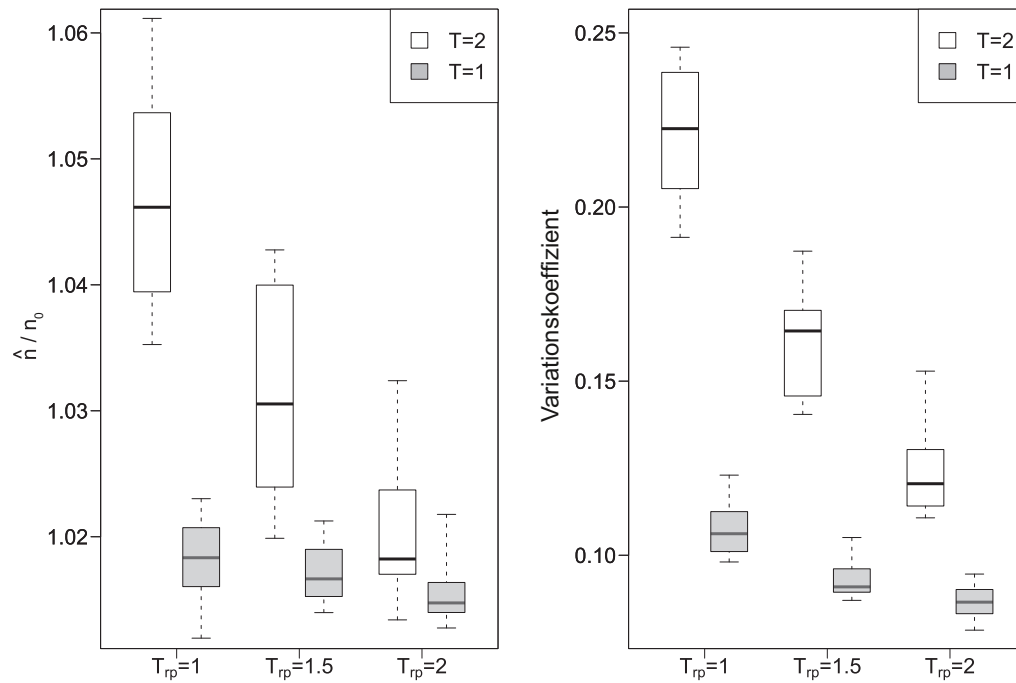


Abbildung 5.3: Zu sehen ist die Interaktion von Studienlänge und Rekrutierungsperiode für das Verhältnis \hat{n}_0/n_0 (a) und den Variationskoeffizienten (b). Die Verteilung ist in Boxplots dargestellt, wobei die Enden der Whiskers das 5. bzw. 95. Perzentil darstellen. Die horizontale Linie ist der Median.

KAPITEL 6

ZUSAMMENFASSUNG UND AUSBLICK

6.1 Zusammenfassung

In den letzten Jahren hat die Bedeutung von adaptiven Studiendesigns mit interner Pilotstudie nach Wittes und Brittain (1990) zunehmend an Bedeutung gewonnen. Für eine Vielzahl an Verteilungen wurden diverse Ansätze zur verblindeten wie auch entblindeten Fallzahlanpassung hergeleitet und publiziert. Das unterstreicht den wachsenden Stellenwert der adaptiven Designs. Einen Überblick verschafft die Arbeit von Friede und Kieser (2006).

In der Literatur sind zwei verschiedenartige Ansätze zur verblindeten Fallzahlanpassung gegeben. Arbeiten von Kieser und Friede (2003), Friede und Kieser (2004) und Friede und Schmidli (2010a,b) vernachlässigen die Gruppenzugehörigkeit und schätzen eine Gesamtstreuung zur Fallzahladaptation, während Gould und Shih (1992) und Cook et al. (2009) einen EM-Algorithmus zur Separation der Behandlungsgruppen und zum Schätzen der Störparameter verwenden. Die EM-Algorithmus-basierende Prozedur von Gould und Shih (1992) wurde für normalverteilte Daten entwickelt, wies jedoch wesentliche Defizite auf (Friede und Kieser, 2002; Waksman, 2007). Ungeachtet dieser Resultate wurde der Ansatz nach Cook et al. (2009) zur Fallzahlanpassung bei Negativ-Binomial-verteilten Daten aufgegriffen. Vor diesem Hintergrund ist in der vorliegenden Dissertation ein Themenkomplex entstanden, der die Prozedur nach Cook et al. (2009) analysiert. Es stellte sich heraus, dass eine Abhängigkeit von Schätzungen und der Wahl des Konvergenzkriteriums besteht, die ebenfalls bei der Normalverteilung beobachtet wurde. Eine sehr große Patientenzahl erweist sich als notwendig, so dass die beobachteten Schätzer zu akzeptablen Ergebnissen führen (vgl. Abschnitt 4.2.3). Infolgedessen kann die EM-Algorithmus-basierende Prozedur für die praktische Anwendung nicht empfohlen werden.

Um Unterschiede zwischen entblindeten und verblindeten Verfahren aufzudecken, wurde ein analytischer Vergleich zwischen den Prozeduren im Poisson-Modell aufgestellt. Anschließend wurden die Resultate in detaillierten Simulationsstudien überprüft. Es wurde eine Regel hergeleitet, mit der

die Überlegenheit der ver- bzw. entblindeten Prozedur gezeigt werden kann. Das zugrunde liegende Entscheidungskriterium bildet die Variabilität der geschätzten Fallzahl. Mit Hilfe dieser Richtlinie wurden vorteilhafte Situationen für beide Prozeduren an klinischen Studien illustriert. Es konnte gezeigt werden, dass eine Überlegenheit der unverblindeten Prozedur eine große Unsicherheit in der Effektstärke impliziert.

Die in der Literatur bestehenden Prozeduren zur Fallzahlanpassung betrachten bisher lediglich konstante Raten (Cook et al., 2009; Friede und Schmidli, 2010a,b). Demgegenüber wurden klinische Szenarien in der multiplen Sklerose und in der chronisch obstruktiven Lungenerkrankung motiviert, in denen diese Annahme nicht geeignet scheint. Aus diesem Grund wurde im Rahmen der Dissertation ein Konzept zur Fallzahlanpassung in klinischen Studien, die zeitvariierende Ereigniszählraten als primären Endpunkt besitzen, entworfen und systematisch analysiert. Die entwickelte Prozedur zur Fallzahlanpassung ist unabhängig von der Wahl der (parametrischen) Baseline-Intensitätsfunktion $\lambda(t)$. Im Kontext des Krankheitsbildes der multiplen Sklerose wurde die jährliche Schubrate beispielhaft durch eine log-lineare Intensität modelliert und in praktisch relevanten Situationen untersucht. Das erweiterte Verfahren weist wünschenswerte Eigenschaften hinsichtlich der Fehlerkontrolle und Studienpower auf. Zusätzlich konnte ein bedeutender praktischer Nutzen gezeigt werden, da die erweiterte Prozedur auf implementierte Routinen in Standardsoftware zurückgreifen kann. Gleichmaßen konnte die bekannte Fallzahlformel nach Keene et al. (2009) und Friede und Schmidli (2010a,b) auf das erweiterte Modell übertragen werden.

6.2 Limitationen der auf Störparameter basierten Anpassung

In Kapitel 4 wurde das Verhalten der unverblindeten wie auch verblindeten störparameterbasierten Fallzahlanpassungsprozeduren bei auftretenden Unsicherheiten der Effektgröße untersucht. In Situationen, in denen eine große Unsicherheit hinsichtlich der Effektgröße besteht, weisen die vorgestellten Anpassungsstrategien keine Robustheit bezüglich der gewünschten Fallzahlverteilung und der Studienpower auf (vgl. Tabelle 4.2). In derartigen Situationen können die auftretenden Limitationen durch Studiendesigns, die eine Effektgröße in der Zwischenauswertung schätzen, umgangen werden. Zum einen eröffnet dieses Vorgehen die Möglichkeit zur Fallzahlanpassung unter Verwendung der beobachteten Effektstärke (effektbasierte Fallzahlanpassung), zum anderen kann anhand dieser Größe ein statistischer Test durchgeführt werden, der über ein vorzeitiges erfolgreiches (erfolgloses) Ende der Studie entscheidet (sequentielle Studiendesigns).

Die ersten Ansätze zur sequentiellen Analyse sind nicht im Gebiet der medizinischen Wissenschaft zu finden. Vielmehr wurden sie im zweiten Weltkrieg zur Einhaltung der militärischen Qualitätsstandards von Abraham Wald und Kollegen entwickelt. Ähnliche Ideen gehen auf George Barnard aus England zurück (Whitehead, 1997; Siegmund, 1985, Kapitel 1). Es gelang jedoch erst in der Nachkriegszeit, dass die Ideen und Methodiken im Buch von Wald (1947) bzw. im Artikel von Barnard (1946) veröffentlicht wurden.

In der medizinischen Anwendung wurden die ersten sequentiellen Ansätze und Studiendesigns von

Armitage (1957, 1960) und Bross (1952, 1958) vorgeschlagen. Das Verfahren von Armitage (1960) verwendet die Annahmen, dass Patientenpaare in die Studie eintreten und auf die Behandlungsarme verteilt werden. Des Weiteren ist eine Zwischenauswertung nach jeder neuen Patienteninformation vorgesehen. Darauf aufbauende gruppensequentielle Methodiken von Pocock (1977) und O'Brian und Fleming (1979) lockern einerseits diese Annahmen, besitzen jedoch andererseits eine Bedingung an die Anzahl der geplanten Zwischenauswertungen sowie eine Charakterisierung der Studiengröße zu diesem Zeitpunkt. Diese Spezifikationen erweisen sich in der Praxis als oftmals restriktiv, so dass auf den alternativen Ansatz, der eine sogenannte „error-spending-function“ verwendet, zurückgegriffen wird (Lan und DeMets, 1983; Kim und DeMets, 1987). Eine flexible Anpassung der Studiengröße und eine parallel stattfindene Zwischenauswertung sind in den beschriebenen Designs nicht möglich (vgl. Restriktionen). Zu diesem Zwecke wurden modifizierte adaptive gruppensequentielle Verfahren entwickelt. Ein Vergleich zwischen der Effizienz der verschiedenen Ansätze wird durch Stallard und Friede (2010); Kapitel 25 bereitgestellt.

6.3 Ausblick und Erweiterungen

Die vorgestellte Prozedur zur Fallzahlanpassung in klinischen Studien, die zeitvariierende Daten als primären Endpunkt besitzen, bietet Raum für folgende Erweiterungen:

1. In der vorliegenden Arbeit wurden proportionale Intensitätsmodelle mit parametrischen Baseline-Intensitätsfunktionen betrachtet. Eine Erweiterung auf nichtparametrische Fragestellungen ist möglich. Eine Fallzahlformel für eine nichtparametrische Teststatistik ist von Matsui (2005) gegeben und kann als Grundlage für weitere Überlegungen verwendet werden. In der Arbeit von Hua und Zhang (2014) wird der Logarithmus der kumulativen Intensitätsfunktion $\Lambda_0(t)$ durch einen B-Spline approximiert. Des Weiteren stellt Hua und Zhang (2014) eine Spline-basierte Maximum-Likelihood Methode und zusätzlich einen Algorithmus zur Berechnung der Schätzer bereit. Eine Schwierigkeit hinsichtlich einer verblindeten Fallzahlanpassung kann bei der Extrapolation der Intensitätsfunktion bei geringer Datenqualität in der Zwischenauswertung auftreten. Bei vorliegenden Überlebenszeitdaten werden ähnliche Probleme der Extrapolation bei unzureichender Datenmenge zum Zeitpunkt der Zwischenauswertung beobachtet und durch Todd et al. (2012) diskutiert. Die untersuchten Methoden könnten auf die vorliegende Fragestellung adaptiert werden (Todd et al., 2012).
2. Das vorgestellte Intensitätsmodell impliziert einen proportionalen Effekt im Studienverlauf. Eine Entwicklung der Prozedur zu Modellen mit zeitvariierenden Effekten wäre wünschenswert. Matsui (2005) und Chen et al. (2013) bieten Ansätze zur Schätzung bei Ereigniszeitdaten mit zeitvariierenden Effekten, wobei Chen et al. (2013) auf Situationen mit vorliegendem „Treatment switch“ fokussiert sind. In Szenarien, in denen der Effekt mit der Zeit variiert, sollten adaptiv sequentielle Designs zur Fallzahlanpassung herangezogen werden.
3. Eine Übertragung der Ergebnisse des proportionalen Intensitätsmodells auf additive Modelle,

in denen der Behandlungseffekt als absolute Risikodifferenz gesehen wird, scheint möglich (Aalen, 1989; Cai und Zeng, 2001; Martinussen et al., 2011).

ANHANG A

WAHRSCHEINLICHKEITSTHEORIE

A.1 Konvergenzen

Definition A.1 (Konvergenz in Wahrscheinlichkeit) Eine Folge von Zufallsvariablen X_n heißt stochastisch konvergent gegen X (in Zeichen $X_n \xrightarrow{\mathcal{P}} X$), wenn

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

gilt.

Definition A.2 (Fast sichere Konvergenz) Eine Folge von Zufallsvariablen X_n konvergiert fast sicher gegen X (in Zeichen $X_n \xrightarrow{f.s.} X$), wenn

$$P(\limsup_{n \rightarrow \infty} \{|X_n - X| > \epsilon\}) = 0 \quad \forall \epsilon > 0$$

gilt.

Definition A.3 (Konvergenz in Verteilung) Eine Folge von Zufallsvariablen X_n konvergiert in Verteilung gegen X (in Zeichen $X_n \xrightarrow{\mathcal{D}} X$), falls

$$P(X_n \leq x) \rightarrow P(X \leq x) \quad \text{für } n \rightarrow \infty$$

an allen Punkten $x \in \mathbb{R}$, an denen $F_X(x) = P(X \leq x)$ stetig ist, gilt.

Satz A.1 (Crámer) Die Abbildung $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ sei auf einer Umgebung von $\mu \in \mathbb{R}^d$ stetig differenzierbar mit $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathbf{X}$. Dann gilt $\sqrt{n}(g(\mathbf{X}_n) - g(\mu)) \xrightarrow{\mathcal{L}} g'(\mu) \cdot \mathbf{X}$. Insbesondere gilt, falls $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}((0), \Sigma)$,

$$\sqrt{n}(g(\mathbf{X}_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, g'(\mu) \Sigma g'(\mu)')$$

BEWEIS. Siehe Kapitel 5 in Ferguson (1996).

□

A.2 Stochastische Prozesse

Definition A.4 (Stochastischer Prozess) Ein stochastischer Prozess ist eine Familie von Zufallsvariablen $\{X(t), t \in I\}$, die über ein und demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ definiert ist.

Definition A.5 (Filtration) Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Eine Familie von σ -Unteralgebren $\{\mathcal{F}_t : t \geq 0\}$ einer σ -Algebra \mathcal{F} heisst Filtration, falls aus $s \leq t$: $\mathcal{F}_s \subset \mathcal{F}_t$ folgt.

Definition A.6 (adaptiert) Ein stochastischer Prozess $\{X(t), t \in I\}$ heisst an die Filtration $\{\mathcal{F}_t : t \geq 0\}$ adaptiert, falls $X(t)$ für jedes $t \geq 0$ ein \mathcal{F}_t -messbarer Prozess ist.

Definition A.7 (Zählprozess) Ein Zählprozess ist ein stochastischer Prozess $\{N(t), t \geq 0\}$, adaptiert an die Filtration $\{\mathcal{F}_t : t \geq 0\}$, für den $N(0) = 0$ und $N(t) < \infty$ f.s. gilt, dessen Pfade stückweise konstant und mit Wahrscheinlichkeit 1 rechtsstetig, bis auf Sprünge an den Unstetigkeitsstellen der Größe +1, sind.

A.3 Elementare, bedingte Wahrscheinlichkeiten

Definition A.8 [Bedingte Wahrscheinlichkeit, vgl. Klenke (2006)] Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $F \in \mathcal{F}$. Dann definieren wir die bedingte Wahrscheinlichkeit gegeben F für jedes $F_1 \in \mathcal{F}$ durch

$$\mathbf{P}[F_1|F] = \begin{cases} \frac{\mathbf{P}[F \cap F_1]}{\mathbf{P}[F]}, & \text{falls } \mathbf{P}[F] > 0 \\ 0, & \text{sonst} \end{cases}.$$

Satz A.2 (Formel von der totalen Wahrscheinlichkeit) Sei I eine höchstens abzählbare Menge und $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $\mathbf{P}[\bigcup_{i \in I} B_i] = 1$. Dann gilt für jede $F \in \mathcal{F}$

$$\mathbf{P}[F] = \sum_{i \in I} \mathbf{P}[F|B_i] \mathbf{P}[B_i].$$

BEWEIS. Vergleiche Klenke (2006). □

Satz A.3 (Bayes'sche Formel) Sei I eine höchstens abzählbare Menge und $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $\mathbf{P}[\bigcup_{i \in I} B_i] = 1$. Dann gilt für jede $F \in \mathcal{F}$ mit $\mathbf{P}[A] > 0$ und jedes $k \in I$

$$\mathbf{P}[B_k|F] = \frac{\mathbf{P}[F|B_k] \mathbf{P}[B_k]}{\sum_{i \in I} \mathbf{P}[F|B_i] \mathbf{P}[B_i]}. \quad (\text{A.1})$$

BEWEIS. Vergleiche Klenke (2006). □

ANHANG B

SIMULATION VON INHOMOGENEN POISSON-PROZESSEN MIT LOG-LINEARER INTENSITÄTSFUNKTION

Anhang B erläutert den Algorithmus zur Simulation von inhomogenen Poisson-Prozessen nach Lewis und Shedler (1976), der in der vorgestellten Simulationsstudie (Abschnitt 5.5) angewandt wird. Im Folgenden wird ein Poisson-Prozess mit der Intensitätsfunktion $\lambda(t) = \exp(\alpha_0 + \alpha_1)t$ in einem fixen Intervall $(0, t_0]$ zugrunde gelegt. Zunächst gilt

$$\Lambda(t) - \Lambda(0) = \int_0^t \lambda(u) du = \begin{cases} \exp(\alpha_0)(\exp(\alpha_1 t) - 1)/\alpha_1, & \alpha_1 \neq 0, \\ \exp(\alpha_0)t & \alpha_1 = 0. \end{cases}$$

Die Anzahl der Ereignisse im Intervall $(0, t_0]$ ist somit durch den Parameter

$$\mu_0 = \exp(\alpha_0)(\exp(\alpha_1 t_0) - 1)/\alpha_1 = \exp(\alpha_0)C/\alpha_1$$

Poisson-verteilt, wobei jedoch nur der interessante Fall $\alpha_1 \neq 0$ betrachtet wird (vgl. Lemma 5.1). Die Zeiten zum Ereignis (times-to-events) sind Ordnungsstatistiken der Verteilung

$$F(t) = \frac{\Lambda(t) - \Lambda(t_0)}{\Lambda(t_0) - \Lambda(0)} = \frac{\exp(\alpha_1 t) - 1}{\exp(\alpha_1 t_0) - 1}, \quad (\text{B.1})$$

wobei $0 \leq t \leq t_0$ $\alpha_1 \neq 0$ ist (vgl. Lewis und Shedler (1976)). Für $\alpha_1 \neq 0$ ist die Gleichung $F(t) = p$ aus (B.1) invertierbar und durch

$$t = F^{-1}(p) = \log(1 + Cp)/\alpha_1$$

gegeben, wobei $0 \leq p \leq 1$. Ein inhomogener Poisson-Prozess kann nun mit Hilfe der Inversions-

methode (z.B. Ahrens und Dieter (1974b)) durch den folgenden Algorithmus generiert werden, der für log-lineare Intensitätsfunktion $\lambda(t) = \exp(\alpha_0 + \alpha_1)t$ beschrieben ist.

Algorithmus ($\alpha_1 \neq 0$):

1. Im ersten Schritt werden n Poisson-verteilte Zufallsvariablen N_i , $i = 1, \dots, n$, mit Rate $\mu_0 = \exp(\alpha_0)C/\alpha_1$ erzeugt. Falls $N_i = 0$ ist, dann wird der Algorithmus hier beendet.
2. Im zweiten Schritt werden n gleichverteilte Zufallsvariablen generiert und geordnet $U_{(1)} \leq \dots \leq U_{(n)}$.
3. Im letzten Schritt wird $\log(1 + CU_{(i)})/\alpha_1$ für $i = 1, \dots, n$, berechnet und die Beobachtungszeiten ergeben sich zu

$$T_1 = \log(1 + CU_{(1)})/\alpha_1, T_2 = \log(1 + CU_{(2)})/\alpha_1, \dots$$

Für geringe Raten μ_0 ist dieser Algorithmus effizient und leicht zu implementieren. Liegen andererseits eine Vielzahl an Ereignissen vor, so gibt es effizientere Methoden, die z. B. durch Ahrens und Dieter (1974a) beschrieben sind. Weitere Verbesserungsmöglichkeiten des vorgestellten Algorithmus und andere Ansätze sind in Lewis und Shedler (1976) gegeben.

LITERATURVERZEICHNIS

- Aalen, O. (1989). A linear regression model for the analysis of life-times. *Statistics in Medicine*, 8:907–925.
- Aban, I. B., Cutter, G. R., und Mavinga, N. (2009). Inferences and power analysis concerning two negative binomial distributions with an application to MRI lesion counts data. *Computational Statistics and Data Analysis*, 53:820–833.
- Ahrens, J. und Dieter, U. (1974a). Computer methods for sampling from gamma, beta, Poisson and binomial distributions. *Computing*, 12:223–246.
- Ahrens, J. und Dieter, U. (1974b). *Non Uniform Random Numbers*. Graz, Austria: Technische Hochschule in Graz.
- Akacha, M. und Benda, N. (2010). The impact of dropouts on the analysis of dose-finding studies with recurrent event data. *Statistics in Medicine*, 29:1635–1646.
- Armitage, P. (1957). Restricted sequential procedures. *Biometrika*, 44:9–26.
- Armitage, P. (1960). *Sequential medical trials*. Oxford: Blackwell.
- Barnard, G. (1946). Sequential tests in industrial statistics. *Journal of the Royal Statistical Society*, 8:1–26.
- Birkett, M. und Day, S. (1994). Internal pilot studies for estimating sample size. *Statistics in Medicine*, 59:2455–2463.
- Bross, I. (1952). Sequential medical plans. *Biometrics*, 8:188–205.
- Bross, I. (1958). Sequential in clinical trials. *Journal of Chronic Diseases*, 8:349–365.
- Cai, J. und Zeng, D. (2001). Additive mixed effect model for clustered failure time data. *Biometrics*, 67:1340–1351.

- Chen, Q., Zeng, D., Akacha, M., Schmidli, H., und Ibrahim, J. (2013). Recurrent events data with treatment switching. *Biometrika*, 100:339–354.
- Chow, S. und Liu, J. (2004). *Design and analysis of clinical trials*. John Wiley & Sons: Hoboken.
- Coffey, C. und Muller, K. (1999). Exact test size and power of a Gaussian error linear model for an internal pilot study. *Statistics in Medicine*, 18:1199–1214.
- Comi, G. and Jeffrey, D., Kappos, L., Montalban, X., Boyko, A., Rocca, M., und Filippi, M. (2012). Placebo-controlled trial of oral Laquinimod for multiple sclerosis. *The New England Journal of Medicine*, 366:1000–1009.
- Committee for Medicinal Products for Human Use (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf. Zugriffen 21 Februar 2014.
- Cook, R., Bergeron, P., Boher, J., und Lie, Y. (2009). Two-stage design of clinical trials involving recurrent events. *Statistics in Medicine*, 28:2617–2638.
- Cook, R. und Lawless, J. (2002). Analysis of repeated events. *Statistical Methods in Medical Research*, 11:141–166.
- Cook, R. und Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer: Berlin.
- Davies, R., Goldberg, A., Forman, S., Pepine, C., Knatterud, G., Geller, N., Sopko, G., Pratt, C., Deanfield, J., und Conti, C. (1997). Asymptomatic cardiac ischemia pilot (ACIP) study two-year follow-up : Outcomes of patients randomized to initial strategies of medical therapy versus revascularization. *Circulation*, 95:2037–2043.
- Day, S. und Altman, D. (2000). Blinding in clinical trials and other studies. *British Medical Journal*, 321:504.
- Dean, C. und Balshaw, R. (1997). Efficiency lost by analyzing counts rather than even times in Poisson and overdispersed Poisson regression models. *Journal of the American Statistical Association*, 92:1387–1398.
- Dempster, A., Laird, N., und Rubin, D. (1977). Maximum likelihood from incomplete data via the EM-algorithm (with discussion). *Journal of the Royal Statistical Society Series B*, 39:1–38.
- Denne, J. und Jennison, C. (1999). Estimating the sample size for a t-test using an internal pilot. *Statistics in Medicine*, 18:1575–1585.
- Deufhard, P. (2011). *Newton methods for nonlinear problems. Affine invariance and adaptive algorithms*. Springer-Verlag: Berlin Heidelberg.

- Ferguson, T. (1996). *A course in large sample theory*. Chapman & Hall: London.
- Food and Drug Administration (2010). Guidance for industry - adaptive design clinical trials for drugs and biologics. draft guidance. <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf>. Zugegriffen 21 Februar 2014.
- Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine*, 317:141–145.
- Friede, T. (1998). Adaptive Fallzahlplanung in klinischen Studien in der t-Test Situation. Master's thesis, Karlsruhe (TH).
- Friede, T., Dunn, J., und Stallard, N. (2008). *Clinical trials in oncology*. In: Missailidis, S.: *Anticancer therapeutics*. John Wiley & Sons: Chichester.
- Friede, T. und Kieser, M. (2002). On the inappropriateness of an EM-algorithm based procedure for blinded sample size re-estimation. *Statistics in Medicine*, 21:165–176.
- Friede, T. und Kieser, M. (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics*, 3:269–279.
- Friede, T. und Kieser, M. (2006). Sample size recalculation in internal pilot study designs: a review. *Biometrical Journal*, 48:537–555.
- Friede, T. und Kieser, M. (2013). Blinded sample size re-estimation in superiority and noninferiority trials: bias versus variance in variance estimation. *Pharmaceutical Statistics*, 12:141–146.
- Friede, T., Mitchell, C., und Müller-Velten, G. (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal*, 49:903–916.
- Friede, T. und Schmidli, H. (2010a). Blinded sample size reestimation with count data: methods and applications in multiple sclerosis. *Statistics in Medicine*, 29:1145–1156.
- Friede, T. und Schmidli, H. (2010b). Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. *Methods of Information in Medicine*, 49:618–624.
- Gould, A. und Shih, W. (1992). Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Communications in Statistics-Theory and Methods*, 21:2833–2853.
- Hilbe, J. (2007). *Negative Binomial Regression*. Cambridge University Press: Cambridge.
- Himmelblau, D. M. (1972). *Applied nonlinear programming*. McGraw-Hill: New York.
- Hua, L. und Zhang, Y. (Zugegriffen 21 Februar 2014). Spline-based semiparametric sieve maximum likelihood method for over-dispersed panel count data. Technical report, can be downloaded from <http://www.public-health.uiowa.edu/biostat/research/documents/technicalreport-huazhang.pdf>.

- ICH (1999). ICH harmonised tripartite guideline E9: statistical principles for clinical trials. *Statistics in Medicine*, 18:1905–1942.
- Inusah, S., Sormani, M., Cofield, S., Aban, I., Musani, S., Srinivasasainagendra, V., und Cutter, G. (2010). Assessing changes in relapse rates in multiple sclerosis. *Multiple Sclerosis*, 16:1414–1421.
- Jenkins, C., Celli, B., Anderson, J., Ferguson, G., Jones, P., Yates, J., und Calverley, P. (2012). Seasonality and determinants of moderate and severe COPD exacerbations in TORCH study. *The European respiratory journal*, 39:1–38.
- Jennison, C. und Turnbull, B. (2000). *Group sequential methods with applications to clinical trials*. Chapman & Hall: London.
- Kappos, L., O Connor, P., Polman, C., Vermersch, P., Wiendl, H., Pace, A., Zhang, A., und Hotermans, C. (2013). Clinical effects of Natalizumab on multiple sclerosis appear early in treatment course. *Journal of Neurology*, 260:1388–1395.
- Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1:305–318.
- Karlis, D. und Xekalaki, E. (2003). Choosing initial values for the em algorithm for finite mixtures. *Computational Statistics and Data Analysis*, 41:577–590.
- Keene, O., Calverley, P., Jones, P., Vestbo, J., und Anderson, J. (2008). Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *European Respiratory Journal*, 32:17i½24.
- Keene, O., Jones, M., Lane, P., und Anderson, J. (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical Statistics*, 6:89–97.
- Keene, O., Vestbo, J., Anderson, J., Calverley, P., Celli, B., Ferguson, G., Jenkins, C., und Jones, P. (2009). Methods for therapeutic trials in COPD: lessons from the TORCH trial. *European Respiratory Journal*, 34:1018–1023.
- Kieser, M. und Friede, T. (2000). Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine*, 19:901–911.
- Kieser, M. und Friede, T. (2003). Simple procedures for blinded sample size adjustment that do not affect the type I error rate. *Statistics in medicine*, 22:3571–3581.
- Kim, K. und DeMets, D. (1987). Design and analysis of group sequential test based on the type I error spending rate function. *Biometrika*, 74:149–154.
- Klenke, A. (2006). *Wahrscheinlichkeitstheorie*. Springer-Verlag: Berlin Heidelberg.

- Konietschke, F., Pauly, M., Schneider, S., Mütze, F., und Friede, T. (2014). On the analysis of treatment effects with time varying over-dispersed count data - asymptotic theory and small sample approximations. (*in Vorbereitung*).
- Lan, K. und DeMets, D. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70:659–663.
- Lawless, J. (1987a). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15:209–225.
- Lawless, J. (1987b). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 82:808–815.
- Lawless, J. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics*, 44:487–498.
- Lewis, P. A. und Shedler, G. (1976). Simulation of nonhomogeneous Poisson processes with log linear rate function. *Biometrika*, 63:501–505.
- Martinussen, T., Scheike, T., und Zucker, D. (2011). The Aalen additive gamma frailty hazards model. *Biometrika*, 98:831–843.
- Matsui, S. (2005). Sample size calculations for comparative clinical trials with over-dispersed poisson process data. *Statistics in Medicine*, 24:1339–1356.
- McCullagh, P. und Nelder, J. (1989). *Generalized Linear Models (2nd ed)*. Chapman & Hall: London.
- McCulloch, C., Searle, S., und J.M., N. (2001). *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics.
- McLachlan, G. und Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics.
- Ng, H. und Tang, M.-L. (2005). Testing the equality of two Poisson means using the rate ratio. *Statistics in Medicine*, 24:955–965.
- Nicholas, R., Straube, S., Schmidli, H., Pfeiffer, S., und Friede, T. (2012). Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: A systematic review and meta-analysis. *Multiple Sclerosis Journal*, 18:1290–1296.
- Nicholas, R., Straube, S., Schmidli, H., Schneider, S., und Friede, T. (2011). Trends in annualized relapse rates in relapsing remitting multiple sclerosis and consequences for clinical trial design. *Multiple Sclerosis Journal*, 17:1211–1217.
- Nordberg, L. (1980). Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models. *Scandinavian Journal of Statistics*, 7:27–32.

- O'Brian, P. und Fleming, T. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35:549–556.
- Pocock, S. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64:191–199.
- Roberts, C. und Torgerson, D. (1998). Randomisation methods in clinical trials. *British Medical Journal*, 317:1301.
- Rogers, W., Bourassa, M., Andrews, T., Bertolet, B., Blumenthal, R., Chaitman, B., Forman, S., Geller, N., Goldberg, A., Habib, G., Masters, R., Moisa, R., Mueller, H., Pearce, D., Pepine, C., Sopko, G., Steingart, R., Stone, P., Knatterud, G., und Conti, C. f. t. A. I. (1995). Asymptomatic cardiac ischemia pilot (ACIP) study: Outcome at 1 year for patients with asymptomatic cardiac ischemia randomized to medical therapy or revascularization. *Journal of the American College of Cardiology*, 26:594–605.
- Rudick, R., Stuart, W., Calabresi, P., Confavreux, C., Galetta, S., Radue, E.-W., Lublin, F., Weinstock-Guttman, B., Wynn, D., Lynn, F., und Panzara, M.A. and Sandrock, A. (2006). Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *New England Journal of Medicine*, 354:911–923.
- Sapatinas, T. (1995). Identifiability of mixtures of power series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, 47:447–459.
- Schneider, S., Schmidli, H., und Friede, T. (2013a). Blinded and unblinded internal pilot study designs for clinical trials with count data. *Biometrical Journal*, 55:617–633.
- Schneider, S., Schmidli, H., und Friede, T. (2013b). Blinded sample size re-estimation for recurrent event data with time trends. *Statistics in Medicine*, 32:5448–5457.
- Schneider, S., Schmidli, H., und Friede, T. (2013c). Robustness of methods for blinded sample size reestimation with overdispersed count data. *Statistics in Medicine*, 32:3623–3635.
- Shreiner, L. (1997). Learning versus confirming in clinical drug development. *Clinical Pharmacology and Therapeutics*, 61:275–291.
- Siegmund, D. (1985). *Sequential Analysis*. Springer-Verlag: New York.
- Sormani, M., Bruzzi, P., Miller, D., Gasperini, C., Barkhof, F., und Filippi, M. (1999). Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *Journal of Neurological Sciences*, 163:74–80.
- Sormani, M., Bruzzi, P., Rovaris, M., Barkhof, F., Comi, G., Miller, D., Cutter, G., und M., F. (2001). Modelling new enhancing MRI lesion counts in multiple sclerosis. *Multiple Sclerosis*, 7:298–304.
- Stallard, N. und Friede, T. (2010). *Efficiency of adaptive designs*. In: Pong, A., and Chow, S.C.: *Handbook of adaptive designs in pharmaceutical and clinical development*. CRC Press.

- Steinvorth, S., Röver, C., Schneider, S., Nicholas, R., Straube, S., und Friede, T. (2013). Temporal trends in baseline characteristics and eligibility criteria in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: A systematic review and meta-regression. *Multiple Sclerosis Journal*, 19:1580–1586.
- Todd, S., Valdés-Márquez, E., und West, J. (2012). A practical comparison of blinded methods for sample size reviews in survival data clinical trials. *Pharmaceutical Statistics*, 11:141–148.
- Torgerson, D. und Roberts, C. (1999). Randomisation methods: concealment. *British Medical Journal*, 319:375–376.
- van den Elskamp, I., Knol, D., Uitdehaag, B., und Barkhof, F. (2009). The distribution of new enhancing lesion counts in multiple sclerosis: further explorations. *Multiple Sclerosis*, 15:42–49.
- Waksman, J. (2007). Assessment of the Gould-Shih procedure for sample size re-estimation. *Pharmaceutical Statistics*, 6:53–65.
- Wald, A. (1947). *Sequential Analysis*. Wiley: New York.
- Wang, Y., Meyerson, L., Tang, Y., und Qian, N. (2009). Statistical methods for the analysis of relapse data in MS clinical trials. *Journal of Neurological Sciences*, 285:206–211.
- Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.
- Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. John Wiley & Sons: New York.
- Wittes, J. und Brittain, E. (1990). The role of internal pilot studies in increasing the efficacy of clinical trials. *Statistics in Medicine*, 9:65–72.
- Wittes, J., Schabenberger, O., Zucker, D., Brittain, E., und Proschan, M. (1999). Internal pilot studies I: type I error rate of the naive t-test. *Statistics in Medicine*, 18:3481–3491.
- Xing, L. und Lee, A. (2005). Sensitivity of test for overdispersion in Poisson regression. *Biometrical Journal*, 47:167–176.
- Zhu, H. und Lakkis, H. (2014). Sample size calculation for comparing two negative binomial rates. *Statistics in Medicine*, 33:376–387.